ToxPredict

OpenTox Online Tutorial

Predict the Toxicity of a Compound with ToxPredict

Apr 20, 2011

OpenTox

Content

- Overview
- ToxPredict demo
- Exercises 1,2,3
  - Define the structure
  - Select models
  - Obtain predictions
  - Browse toxicity data
- Additional slides (optional)
  - ToxPredict behind the scenes

OpenTox

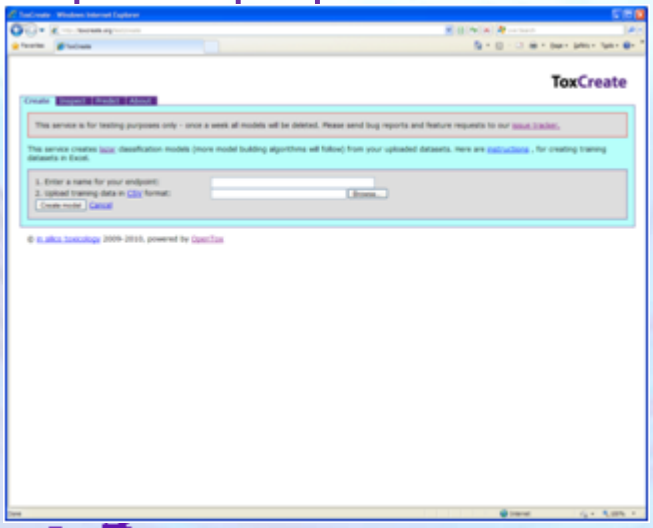
OpenTox framework

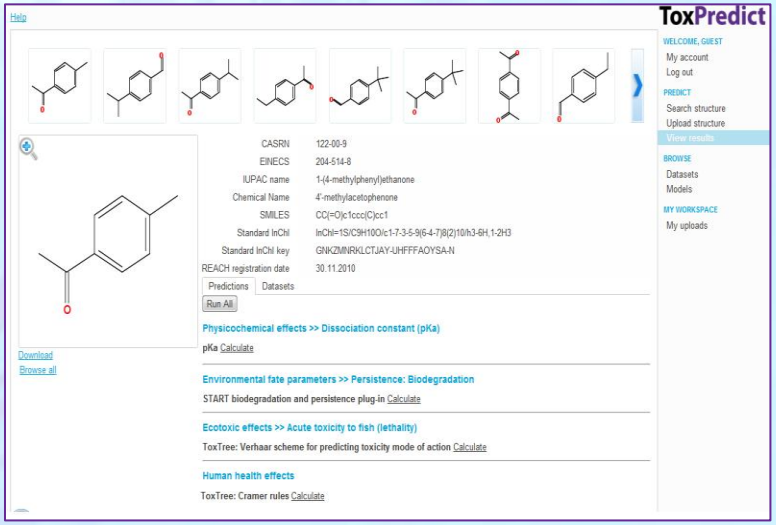
- Distributed Web Services for predictive toxicology
- Several types of Web Services (using REST Web service technology)
- Service types corresponds to following building blocks :
  - Chemical compound;
  - Dataset of chemical compounds and their properties (calculated or measured)
  - Algorithm (descriptor calculation, regression, classification, structural alerts, QC, etc.)
  - Predictive model
  - Report, validation, applicability domain, etc.
- Every object (compound, dataset, algorithm, model, etc.) has an unique web address (e.g. http://myhost.com/model/bestpredictivemodel)
- These objects can be created, read, deleted, and updated
- Every object has an RDF (W3C Resource Description Framework) representation, defined in OpenTox ontology (opentox.owl)

OpenTox

Demo applications

- Two end user oriented demo applications, making use of OpenTox webservices, have been developed, deployed and are available for testing – **http://toxcreate.org** and **http://toxpredict.org** ;
- ToxCreate creates models from user supplied datasets;
- ToxPredict uses existing OpenTox models to estimate chemical compound properties



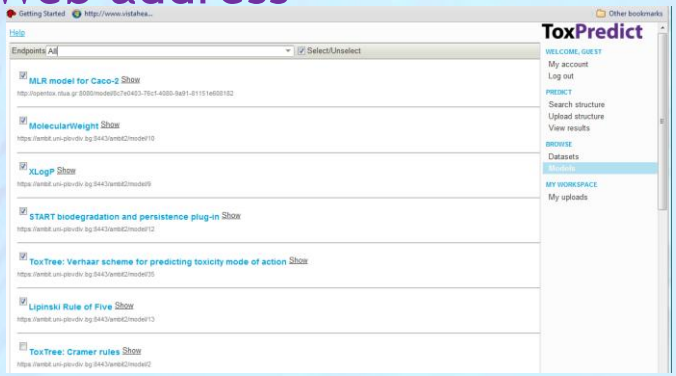


OpenTox

ToxPredict at http://toxpredict.org

ToxPredict estimates the chemical hazard of chemical structures. It relies on **OpenTox API-v1.1** compliant RESTful webservices. Users can either search the OpenTox prototype database, which includes currently quality labelled data for **~150,000 chemicals**, grouped in number of datasets, or upload their own chemical structure data. ToxPredict provides access to **16 ready to use models**, addressing **14 different endpoints** (and growing);

- ToxPredict uses the following OpenTox webservices: **Compound**, **Feature**, **Dataset**, **Algorithm**, **Model** , **Task** , **Ontology**, **Validation**;
- Models, available via OpenTox API can be easily integrated in ToxPredict by just publishing its Web address



OpenTox

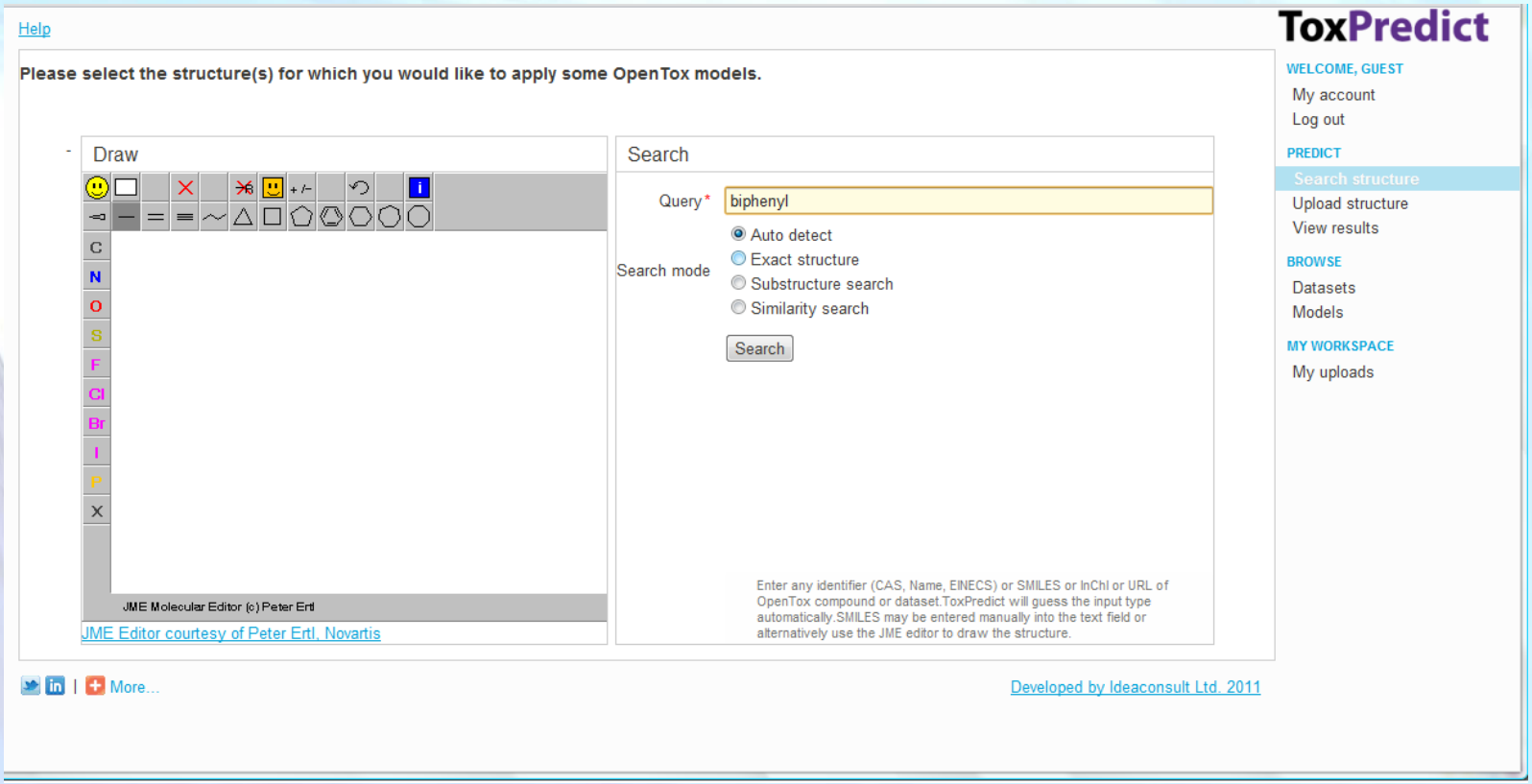
Exercise 1

- Search for a structure by chemical name
- Run models and obtain predictions
- Explore toxicity data for the selected compound
- Browse a dataset with toxicity data.

OpenTox

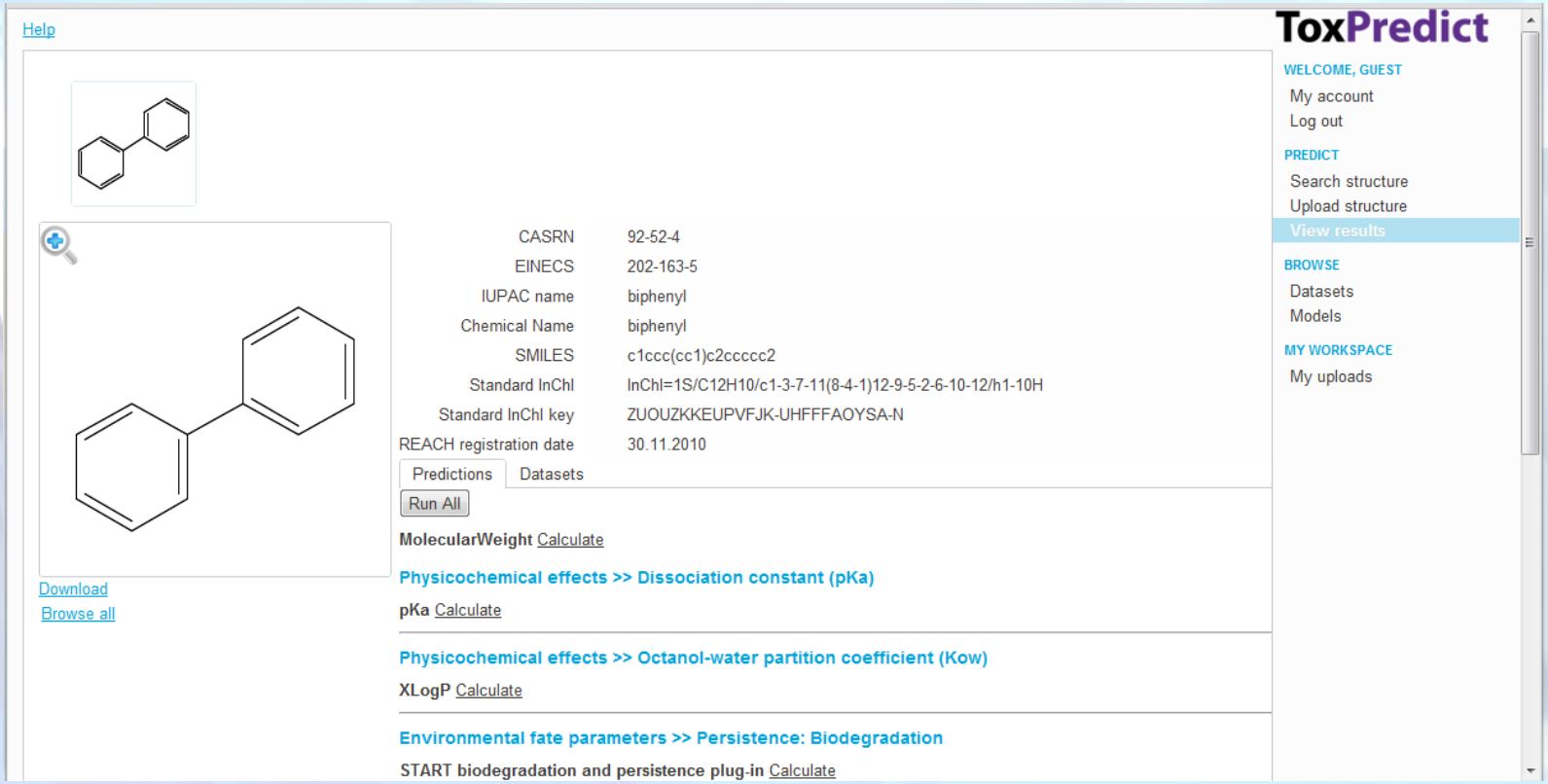
Define the structure

Select “Search Structure” .
Enter “Biphenyl” in the query text box. Press “Search” button.



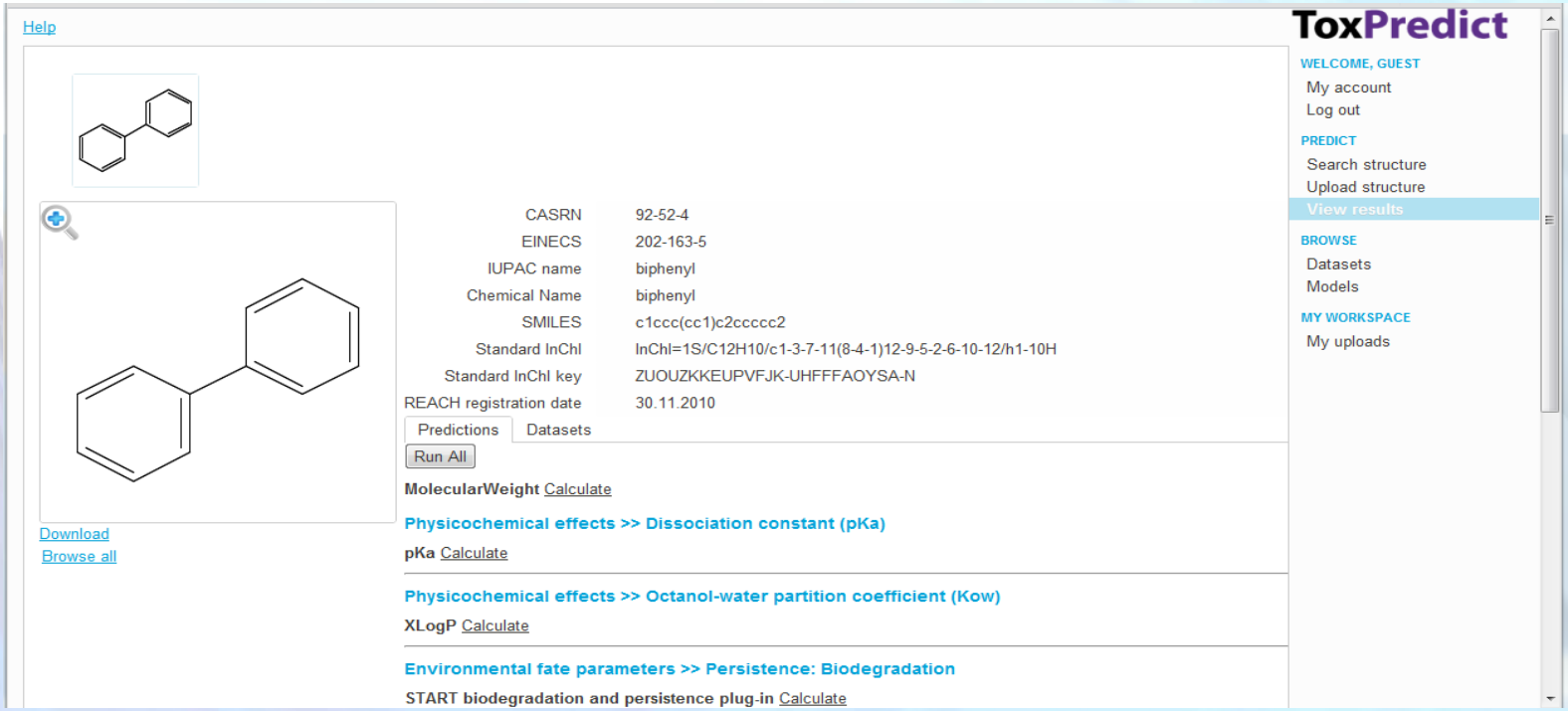
View the results and run models

The available models are listed under the “Predictions” tab.
Click on the “Calculate” link next to each model to obtain predictions.



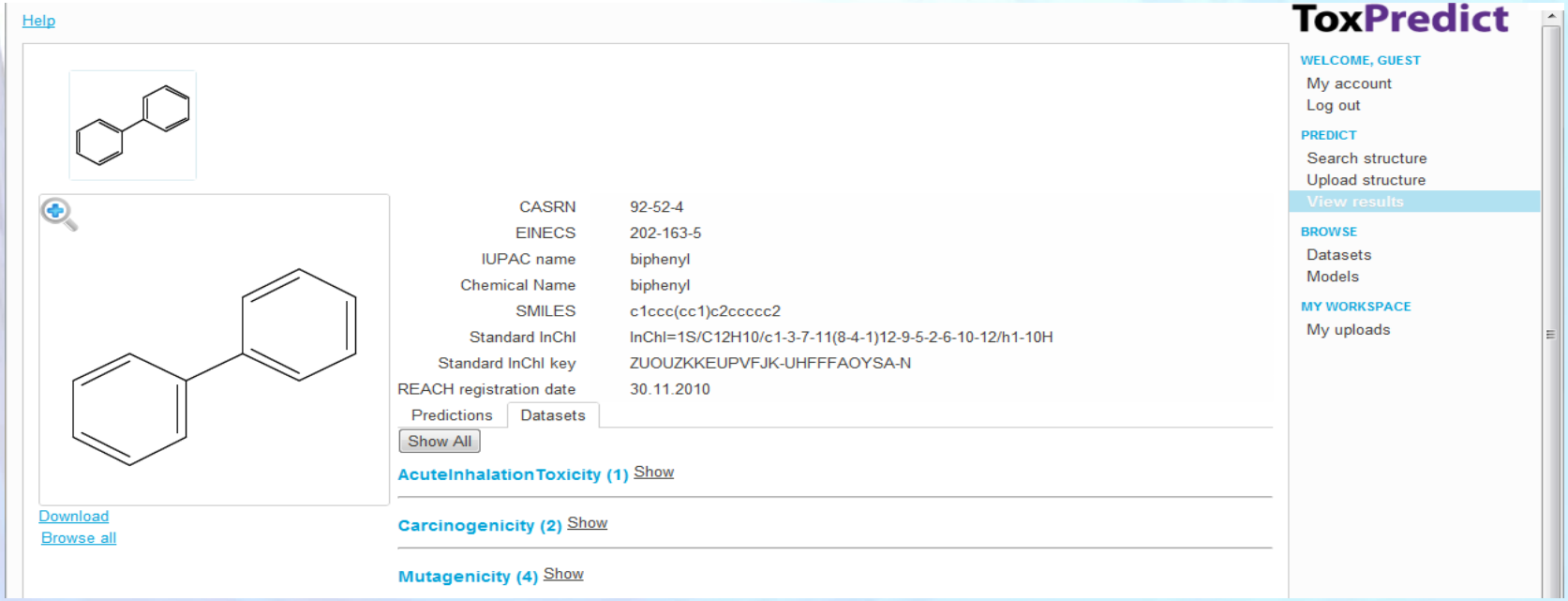
View the results and run models

The available models are listed under the “Predictions” tab.
Click on the “Calculate” link next to each model to obtain predictions.
Chemical structure and prediction results can be downloaded via the “Download” link.



Browse toxicity data for the selected compound

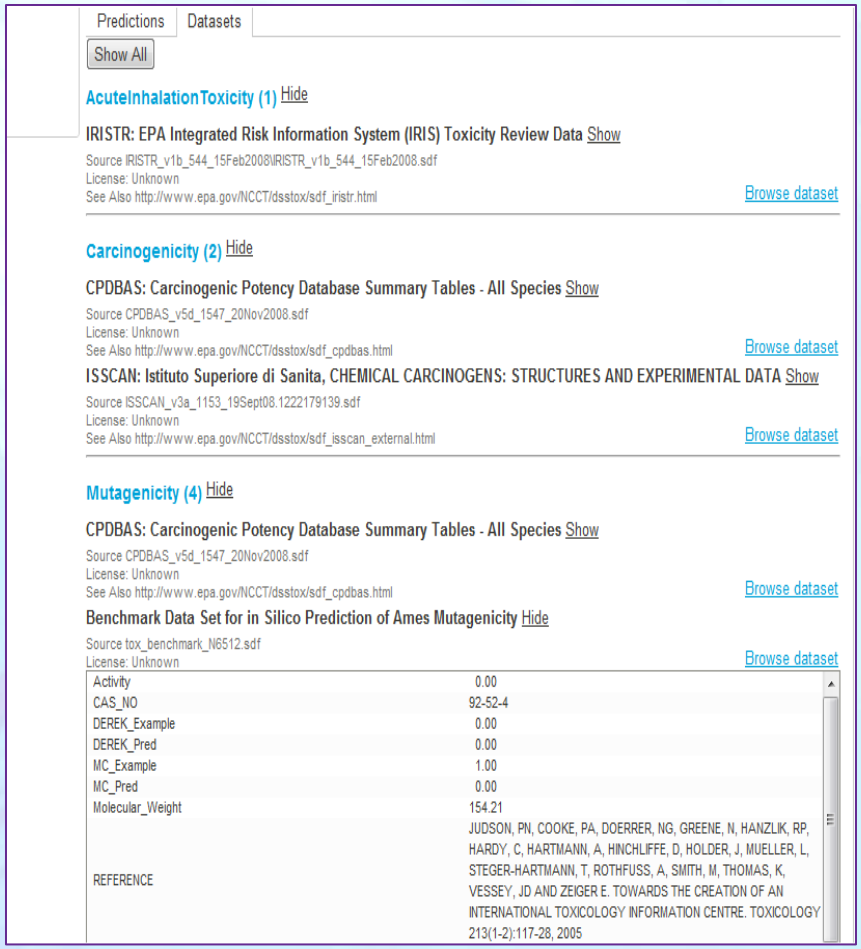
The toxicity data for the selected compound could be accessed at the “Datasets” tab. The top level categories shows the endpoint name and the number of datasets (in brackets) containing toxicity information for the selected compounds. Click the “Show” link to view the list of the datasets.



OpenTox

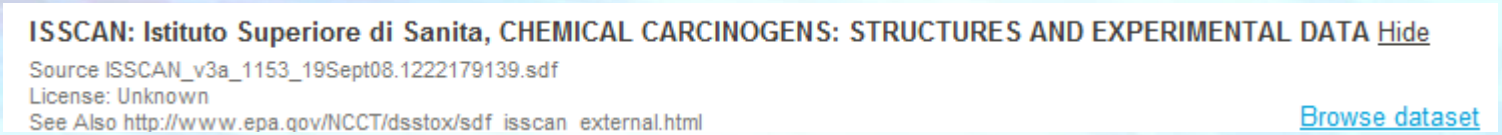
Browse toxicity data for the selected compound

- Pressing the "Show" link next to the endpoint name list the available datasets.
- The "Show" link next to the dataset title displays the content for the selected compound.
- To browse the entire dataset, use "Browse dataset" link. Note this will bring you to another page, but you could access the current page either via browser "Back" button, or the ToxPredict "View Results" menu on the right.

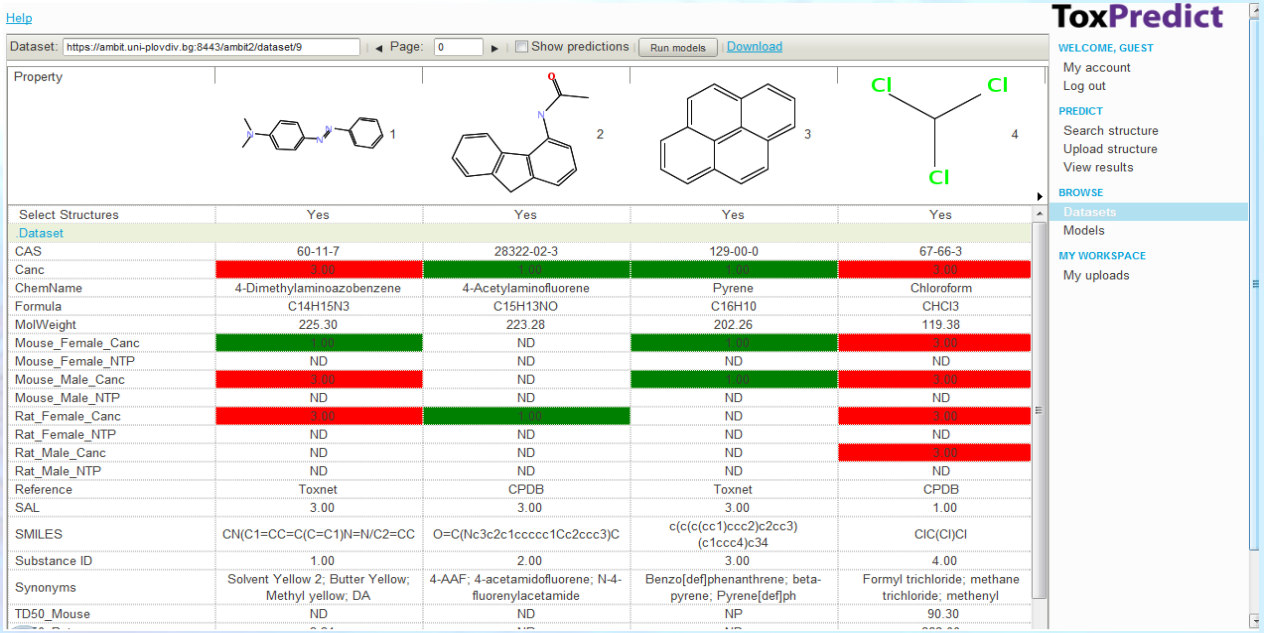


OpenTox

Browse a toxicity dataset



- Click the "Browse dataset" link next to the ISSCAN dataset. This will show the "Datasets" page.



Property	1	2	3	4
Select Structures	Yes	Yes	Yes	Yes
Dataset				
CAS	60-11-7	28322-02-3	129-00-0	67-66-3
Canc	3.00	1.00	1.00	3.00
ChemName	4-Dimethylaminoazobenzene	4-Acetylaminofluorene	Pyrene	Chloroform
Formula	C14H15N3	C15H13NO	C16H10	CHCl3
MolWeight	225.30	223.28	202.26	119.38
Mouse_Female_Canc	1.00	ND	1.00	3.00
Mouse_Female_NTP	ND	ND	ND	ND
Mouse_Male_Canc	3.00	ND	1.00	3.00
Mouse_Male_NTP	ND	ND	ND	ND
Rat_Female_Canc	3.00	1.00	ND	3.00
Rat_Female_NTP	ND	ND	ND	ND
Rat_Male_Canc	ND	ND	ND	3.00
Rat_Male_NTP	ND	ND	ND	ND
Reference	Toxnet	CPDB	Toxnet	CPDB
SAL	3.00	3.00	3.00	1.00
SMILES	CN(C1=CC=C(C=C1)N=N/C2=CC	O=C(Nc3c2c1ccccc1Cc2ccc3)C	c(c(c(cc1)ccc2)c2cc3)(c1ccc4)c34	ClC(Cl)Cl
Substance ID	1.00	2.00	3.00	4.00
Synonyms	Solvent Yellow 2; Butter Yellow; Methyl yellow; DA	4-AAF; 4-acetamidofluorene; N-4-fluorenylacetamide	Benzo[def]phenanthrene; beta-pyrene; Pyrene[def]ph	Formyl trichloride; methane trichloride; methenyl
TD50_Mouse	ND	ND	NP	90.30

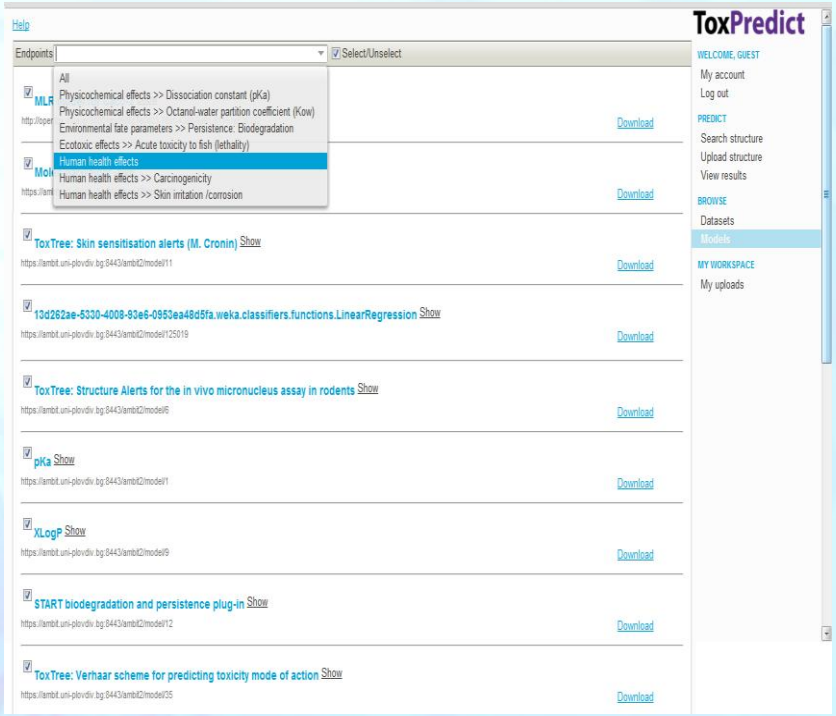
Exercise 2

- Select subsets of models to be applied
- Draw a structure
- Search for similar compounds
- View the results and obtain predictions.
- Explore toxicity data for the selected compounds
- Run model predictions in batch mode.
- Browse the results

OpenTox

Select a subset of models

- **Select "Models" from ToxPredict menu on the right.**
- The list of models appears. Only selected models will be displayed/used when running predictions on other ToxPredict pages.
- Select/Unselect models one by one or by filtered subsets.
  - Use the Endpoints drop down box to filter models by endpoint.
  - The "Select/Unselect" checkbox on the toolbar is applied only on currently visible models.
- Click the "Show" link next to each model to view more information about a model

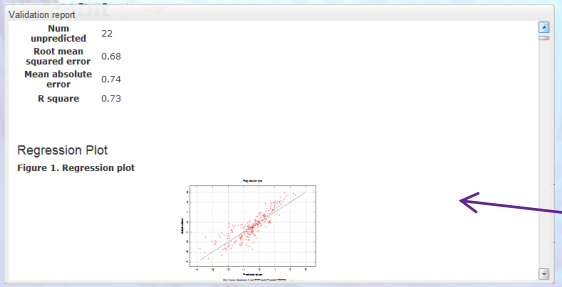


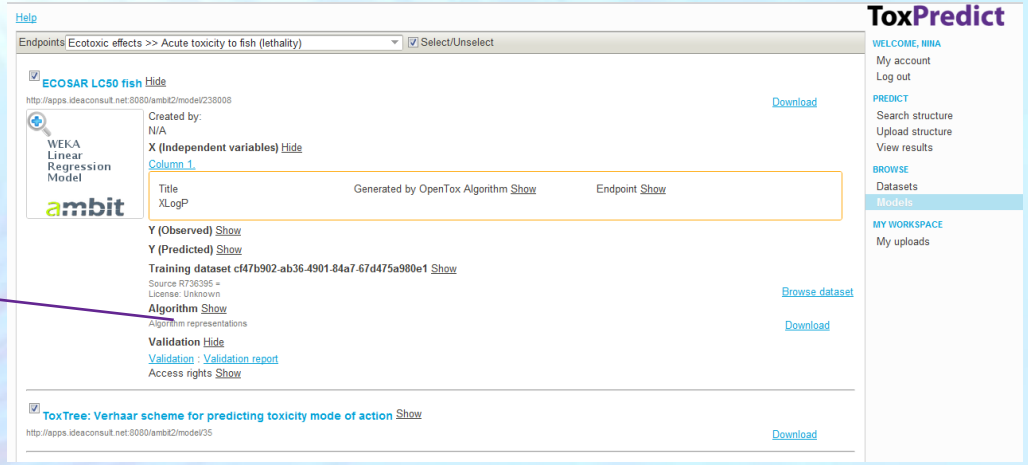
When ready, click on "Search structure" menu on the right.

Model details

Model details includes information about the algorithm, used to derive the model, independent and dependent variables, training dataset, access rights , and links to validation performed and resulted validation reports.

Models are OpenTox web services, derived by learning algorithms, also OpenTox web services.
Validations and validation reports are created by the OpenTox Validation service.



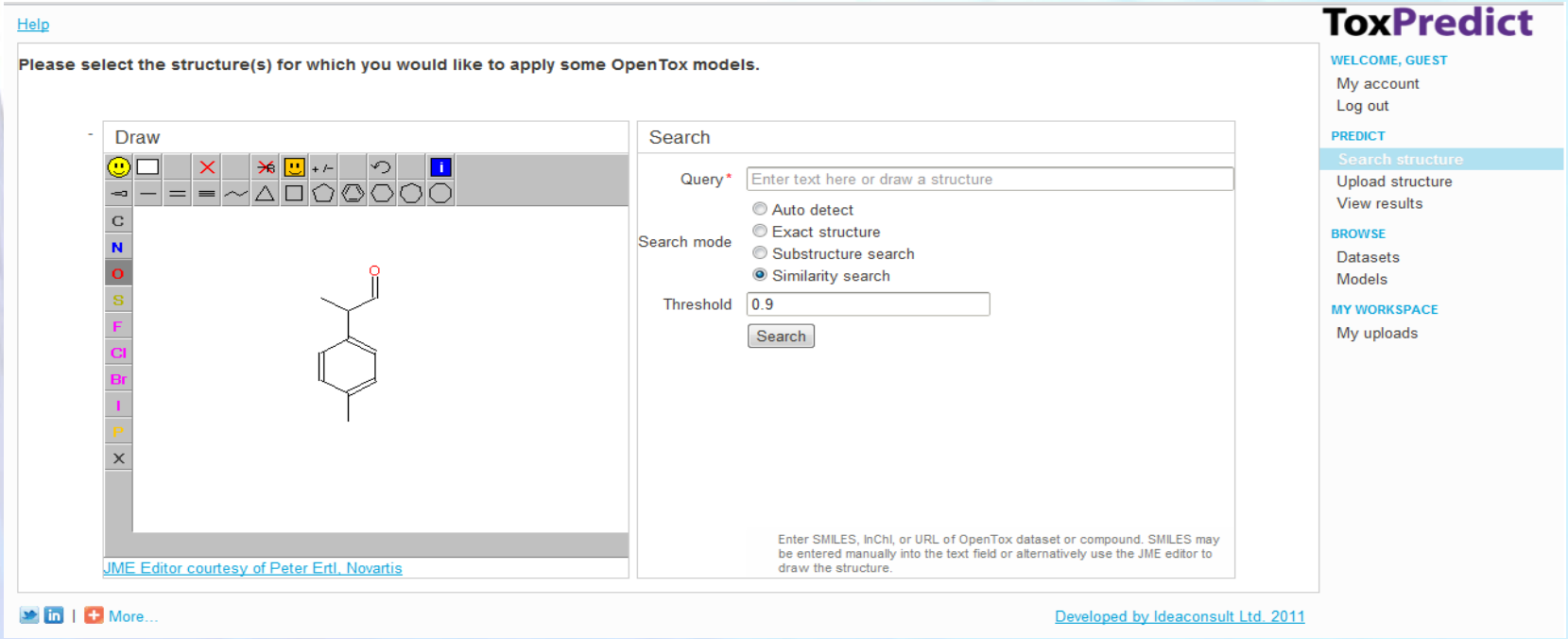


When ready, click on "Search structure" menu on the right.

Define the structure and search mode

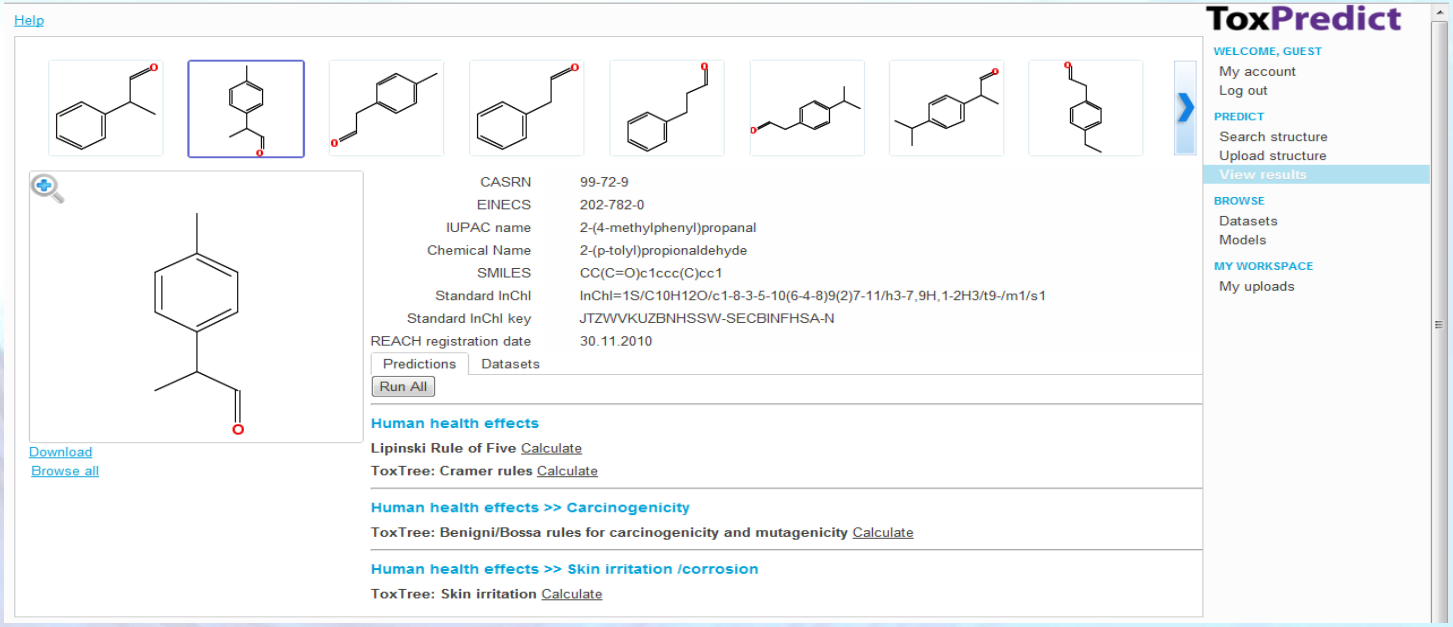
Select "Search Structure" from ToxPredict menu on the right.

- Draw the structure of *2-(4-methyl phenyl) propanal*.
- Select "Similarity search" as a search mode
- Press the "Search" button.



View the results and run models

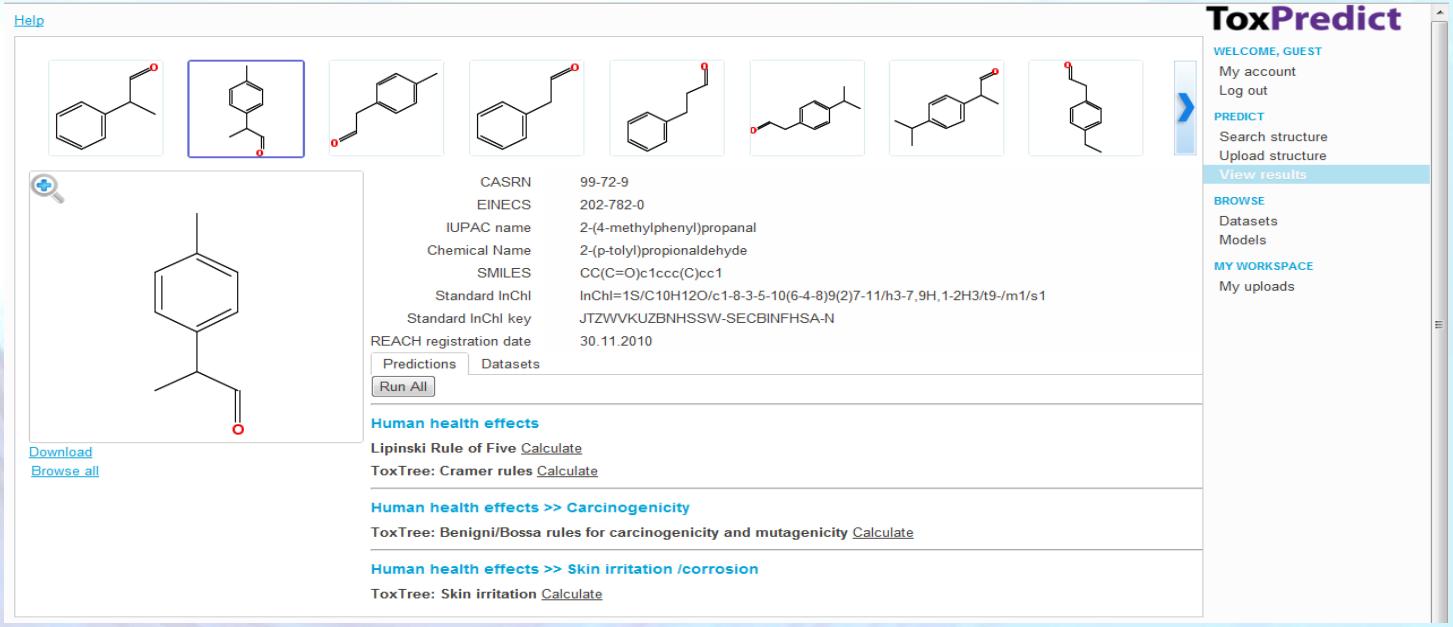
The selected models are listed under the “Predictions” tab.
Click on the “Calculate” link next to each model to obtain predictions.
Select the “Dataset” tab to verify if there is toxicity data for a compound.



As an alternative to inspecting structures one by one, click on the “Browse All” link on the left (just below the selected compound diagram)

View the results and run models

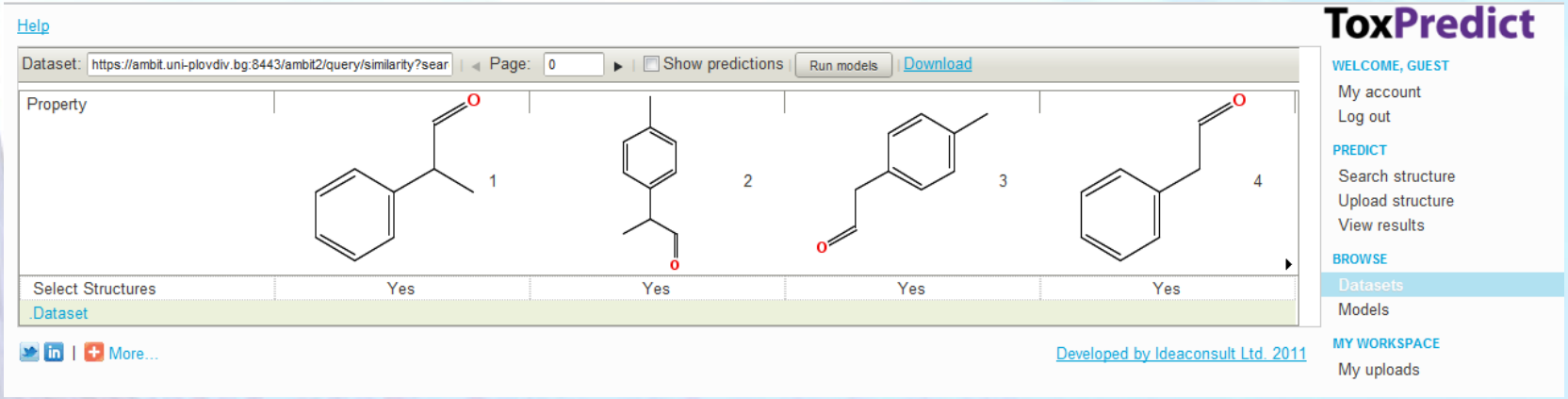
The selected models are listed under the “Predictions” tab.
Click on the “Calculate” link next to each model to obtain predictions.
Select the “Dataset” tab to verify if there is toxicity data for a compound.



As an alternative to inspecting structures one by one, click on the “Browse All” link on the left (just below the selected compound diagram)

View the results and run models (batch mode)

The similarity search hits are displayed in a table, where columns are compounds, and rows are properties or actions. Page navigation controls are available on the toolbar on the top.

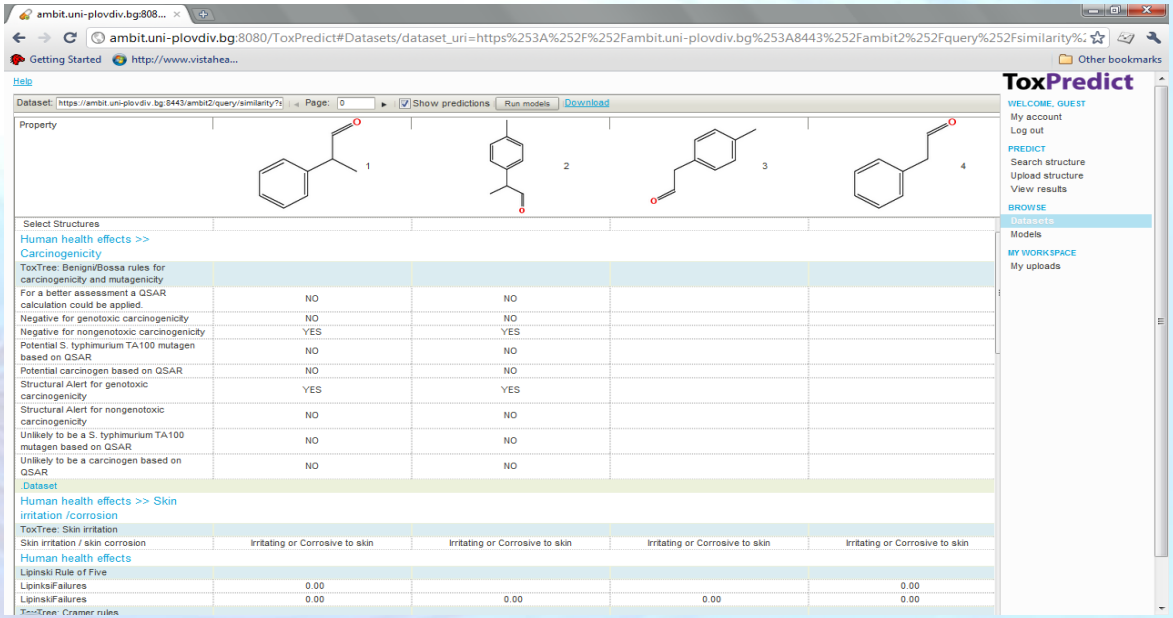


To display model predictions , select “Show predictions” checkbox on the toolbar. This will add as many rows, as necessary to represent the results of applying the previously selected models.

View the results and run models (batch mode)

The similarity search hits are displayed in a table, where columns are compounds, and rows are properties or actions. Page navigation controls are available on the toolbar on the top.

If not all compounds have prediction results, click "Run models" button to run predictions . The results can be downloaded via the "Download" link.



OpenTox

Exercise 3

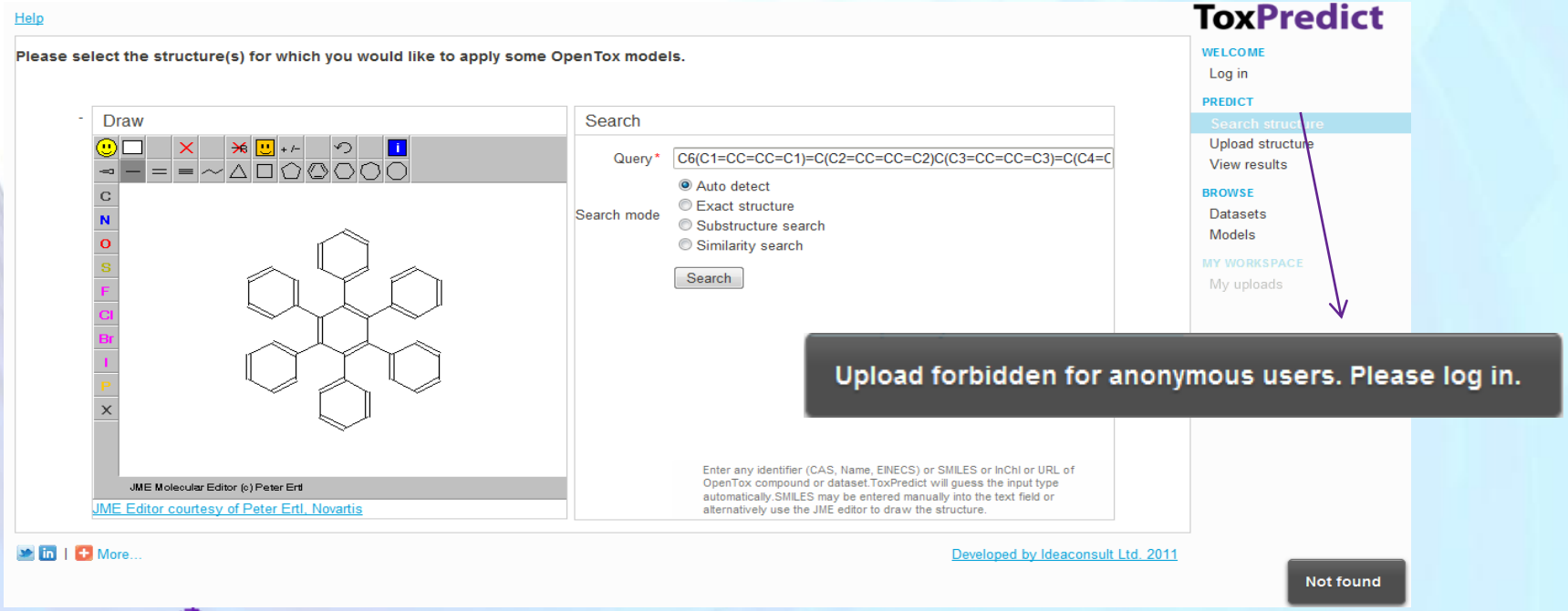
- Upload new chemical structure
- Example file

  http://opentox.org/data/documents/development/tutorialfiles/toxpredict/ToxPredictExercise3/view
- Run models and obtain predictions
- Explore “My uploads” page

Dealing with “Structure not found”

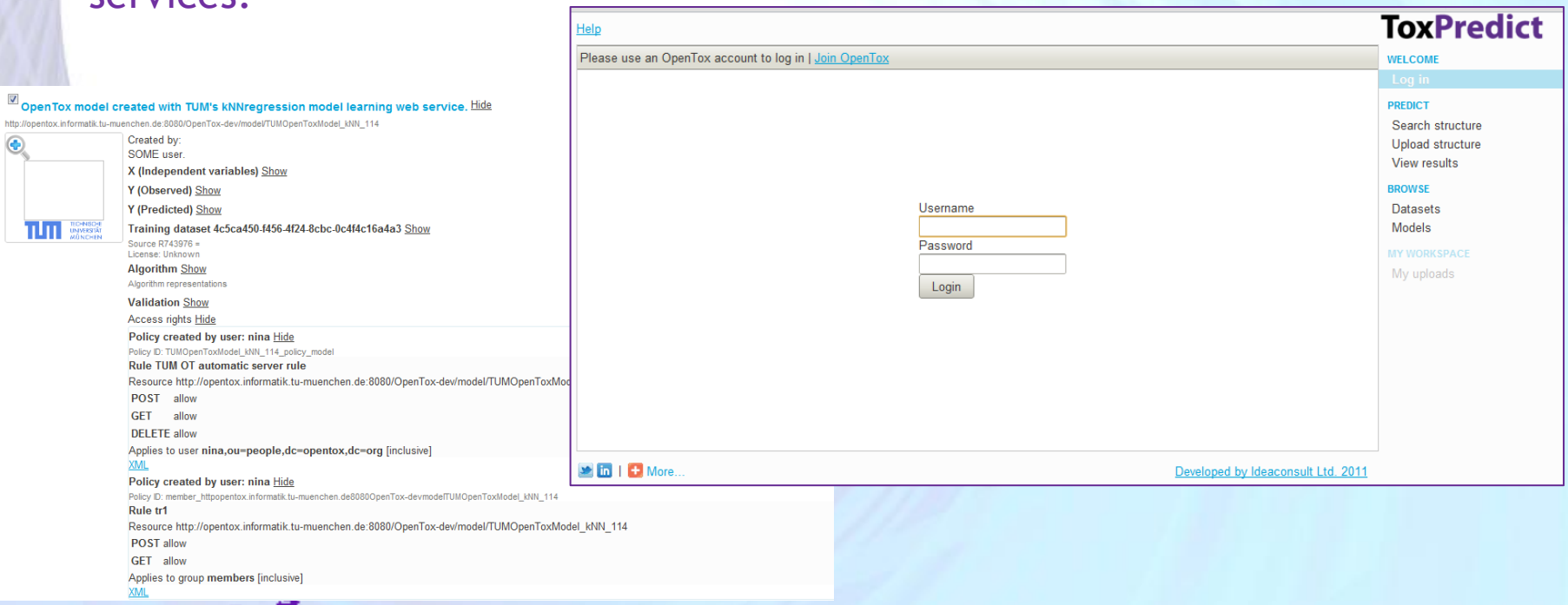
The “Search structure page” queries the OpenTox dataset service. If the chemical structure is not found in the database, behind the service, it will return “Not found” message.

However, any **registered user** could upload chemical structures and data via “Upload structure” page.



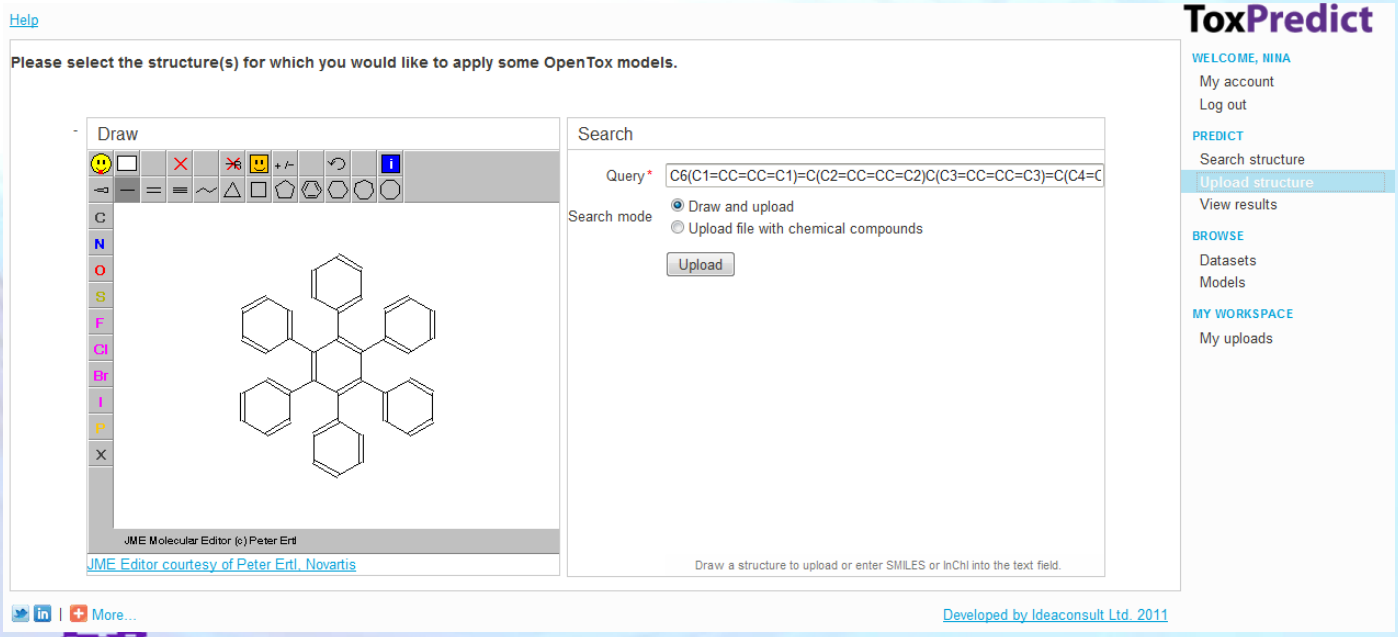
Login in

OpenTox REST web services include both publicly accessible and protected resources. Both data and computational resources could be protected. OpenAM service http://forgerock.com/openam.html is used for this purpose. The same users as registered on www.opentox.org site is used across all web services.



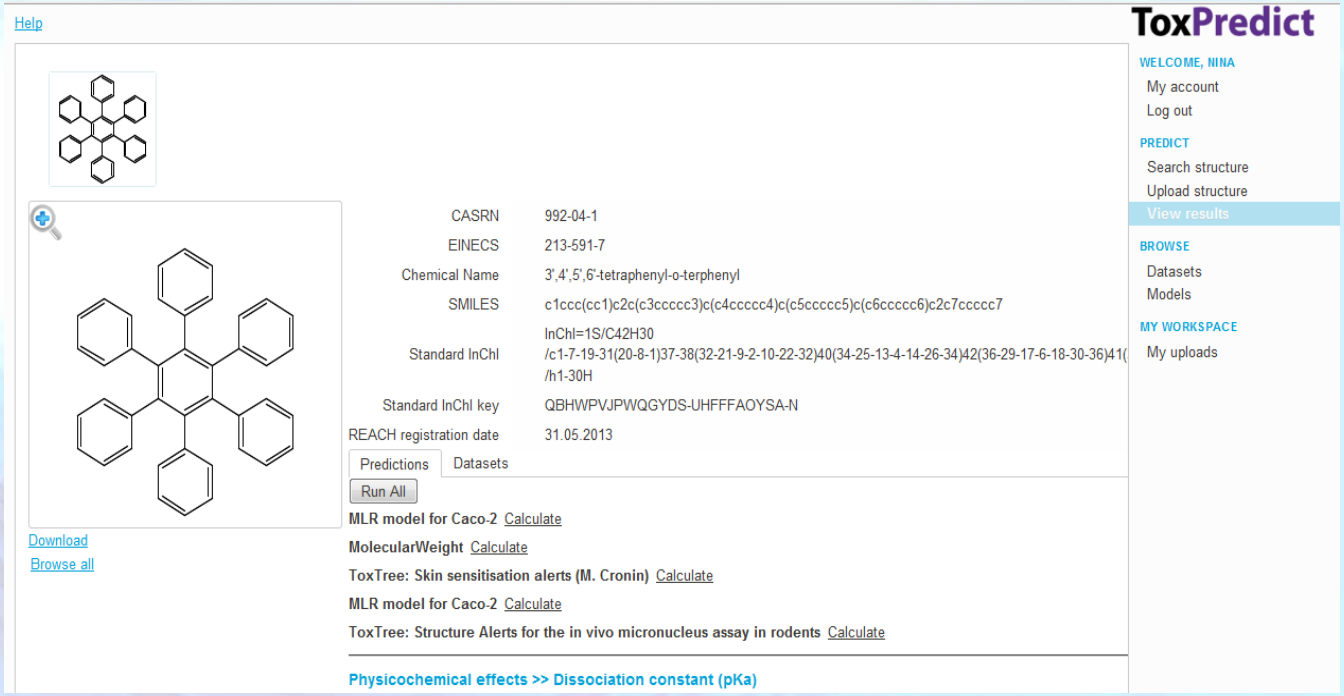
Upload structures

Switch to “Upload structure” page. The structure, drawn in “Search structure” page is retained, but the options next to it allow to upload the structure drawn or a new file.
Click Upload button when ready.



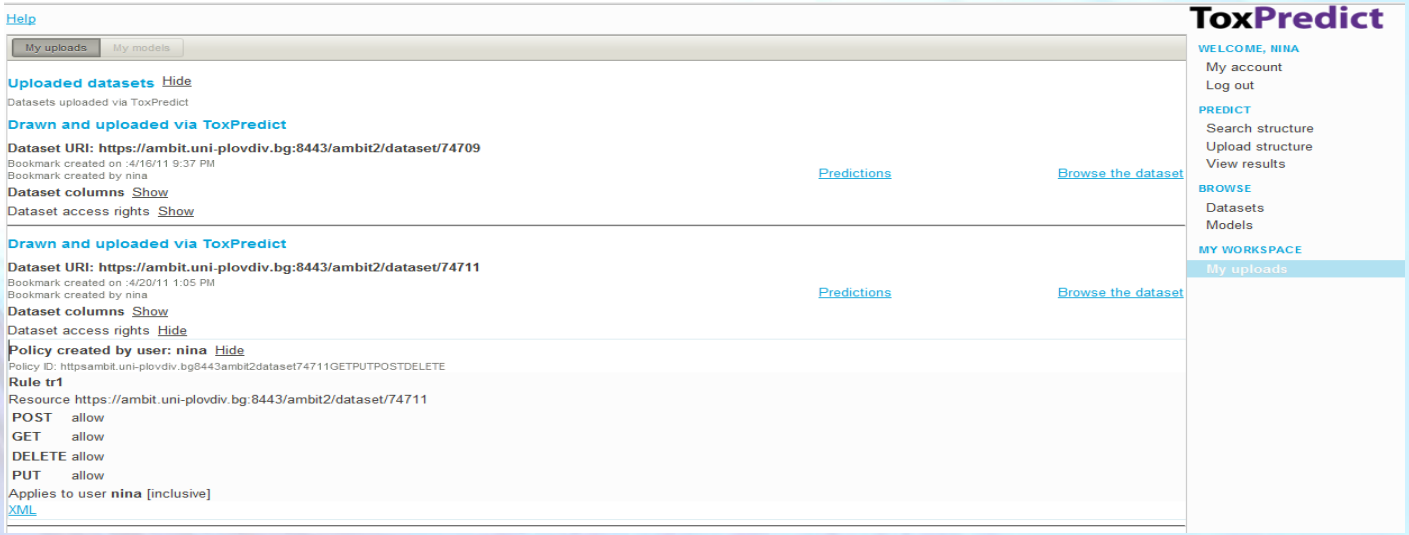
View uploaded structures, apply models

Upon upload, the structure(s) will be displayed in the “View results” page.



“My uploads” page

Lists the files and structures uploaded by the currently logged-in user. If the dataset service allows protected datasets, the assigned access rights will allow access only to the creator of the resource. These could be changed later , but only by the creator.



The dataset service, used in the tutorial , doesn’t allow protected resources. All structures uploaded today will become publicly accessible!

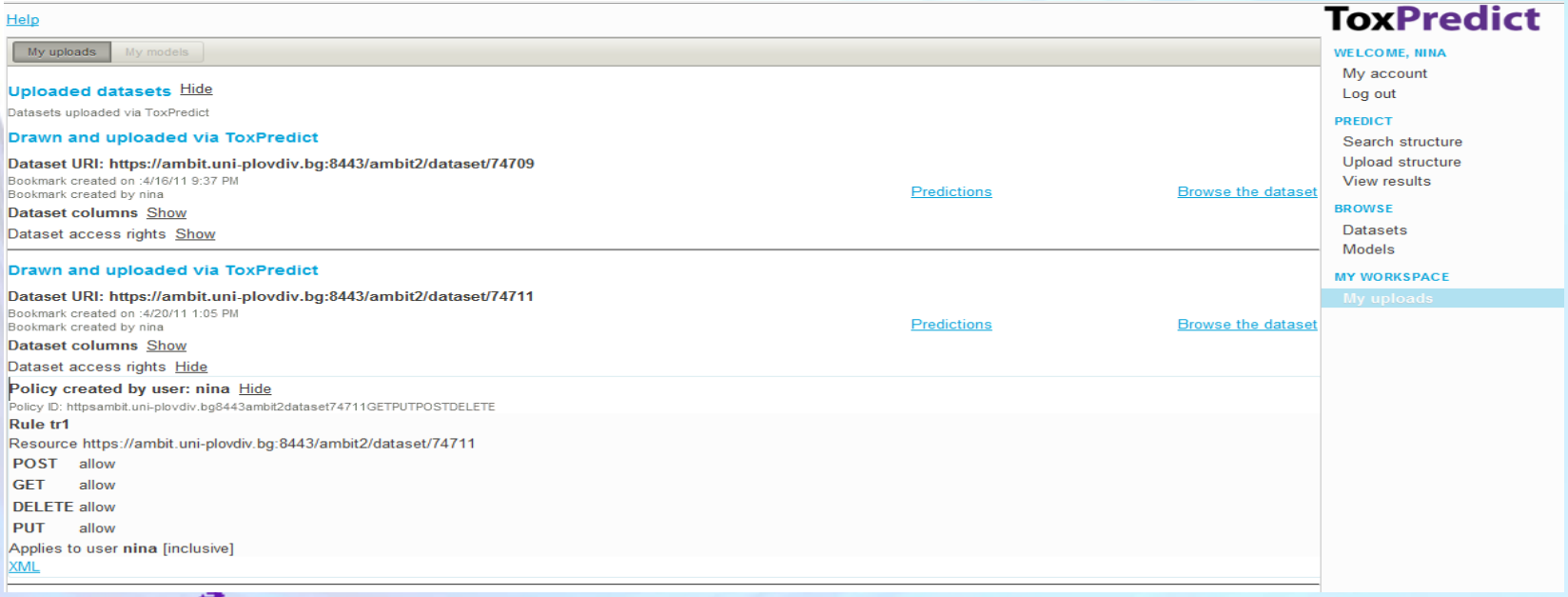
“My uploads” page

The uploaded resources will be retained after log on. They could be reused on subsequent logon.

- Click on “Predictions” link to open the datasets into “View Results” page

  or

- Click on “Browse dataset” to view the structures and data in table mode.



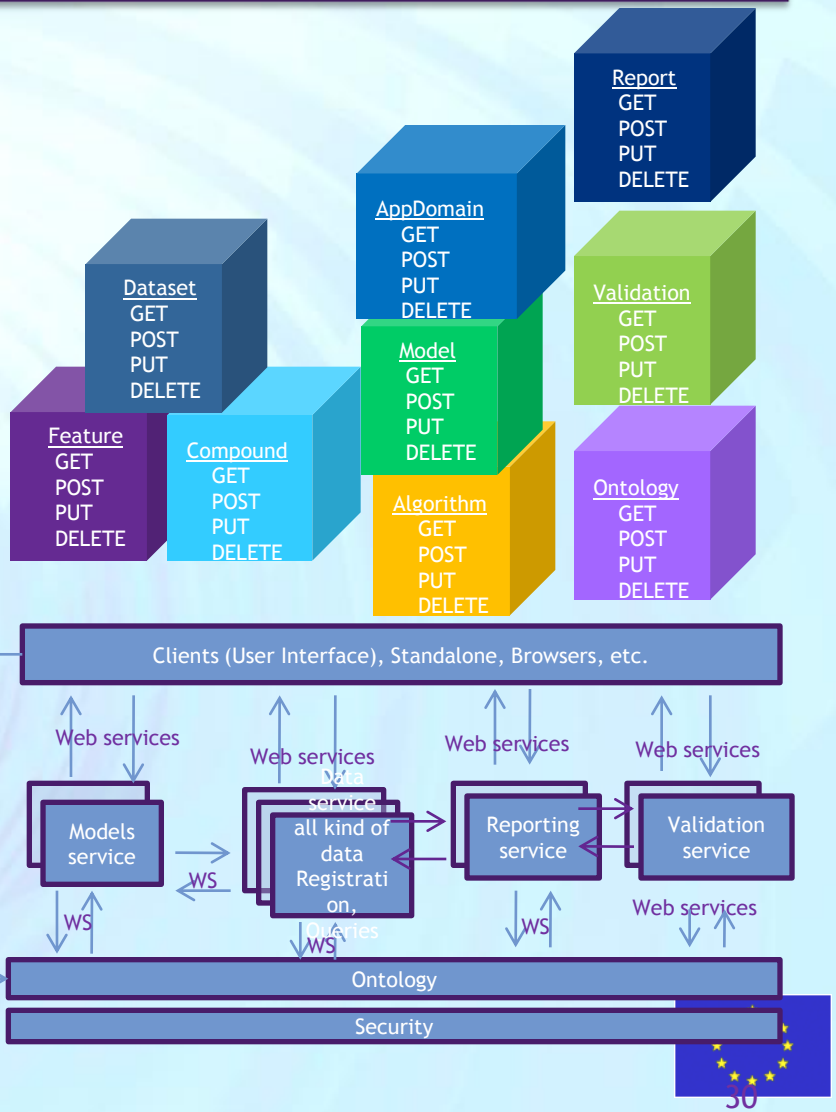
End of exercises.

Behind the scenes

The following slides present a more in-depth overview of ToxPredict's workflow and provide some more details on its interactions with OpenTox webservices which are taking place behind the scenes and without requiring any end-user intervention;

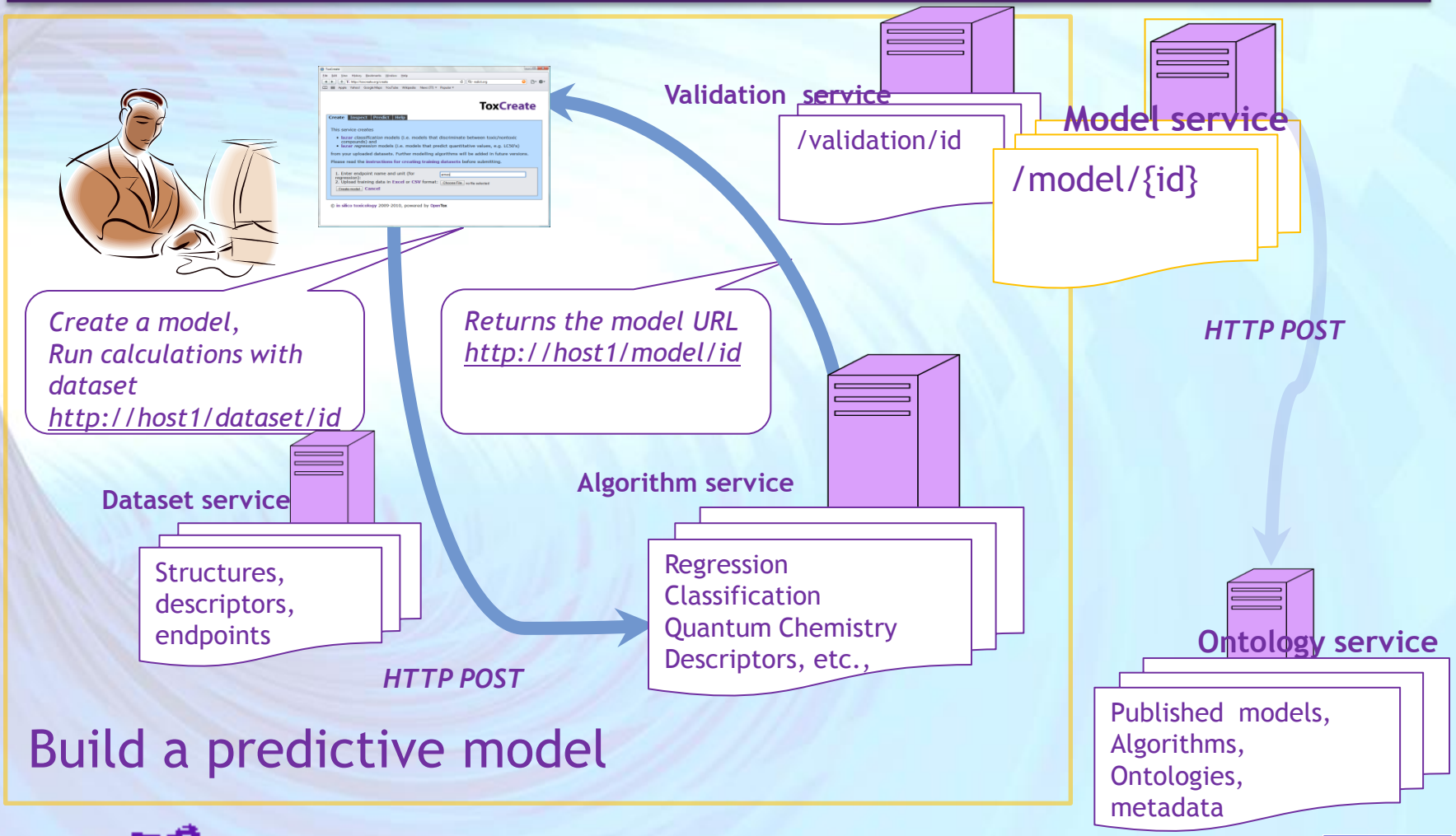
OpenTox framework

- OpenTox API
  - The way applications talk to each other
  - The way developers talk to applications
  - http://opentox.org/dev/apis/api-1.1
- The basic building blocks:
  - data, chemical structures, algorithms and models.
- Functionality offered
  - build models,
  - apply models,
  - validate models,
  - access and query data in various ways.
- Technologies
  - REST style web services
  - RDF for description of resources
  - Links to existing and newly developed ontologies (mainly to describe metadata) about resources



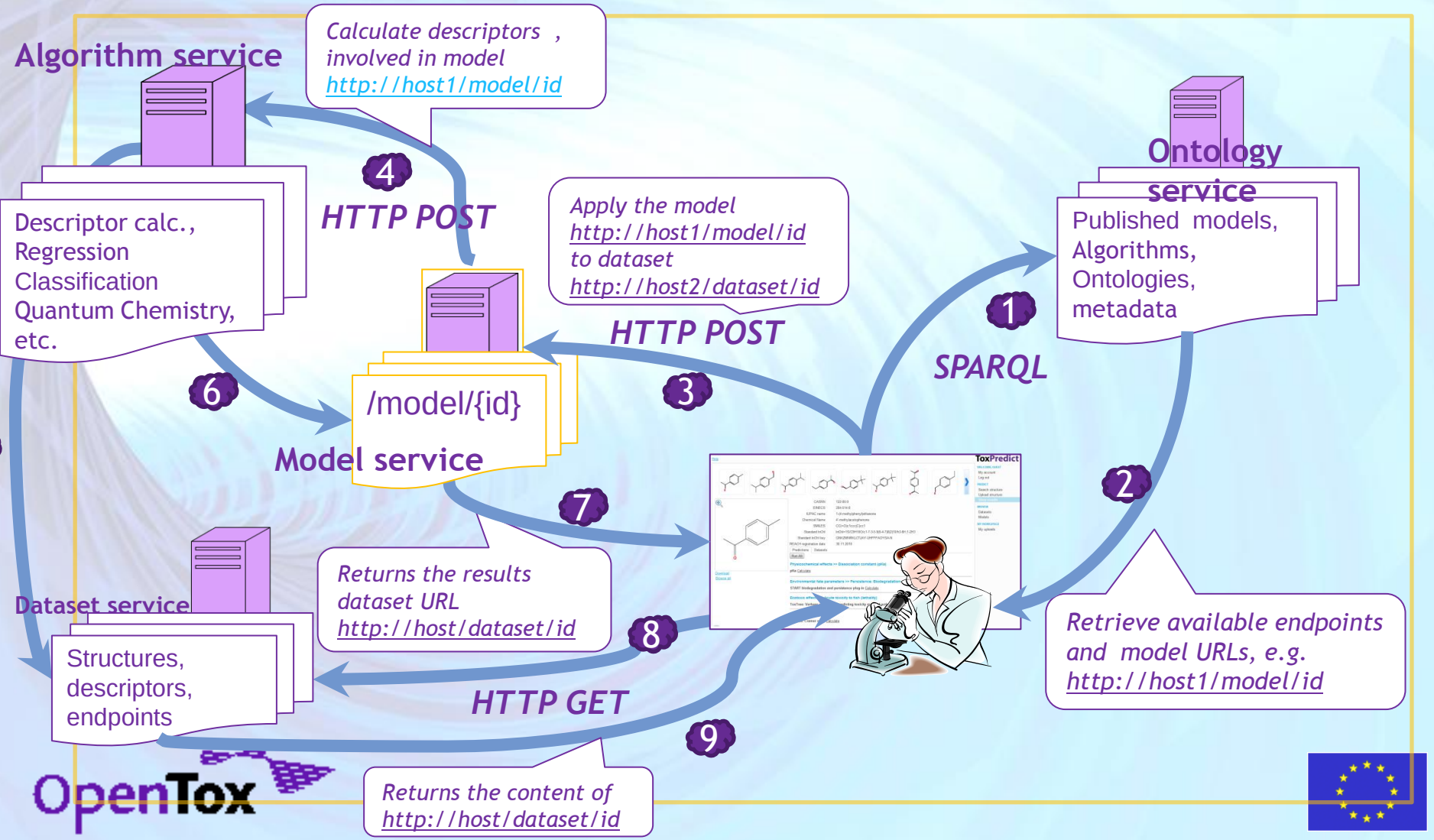
OpenTox

Build a predictive model (ToxCreate)



OpenTox

Apply predictive models (ToxPredict)



ToxPredict: Step 1 (Select structure(s))

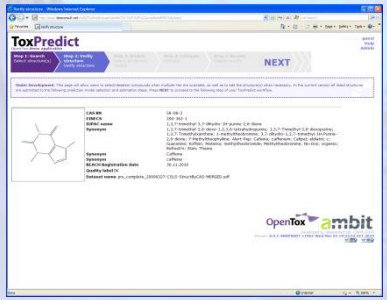
Find structure by name, registry number, SMILES, InChI, structure, substructure, similarity…

ToxPredict Web Application

OT Dataset API ***HTTP GET***

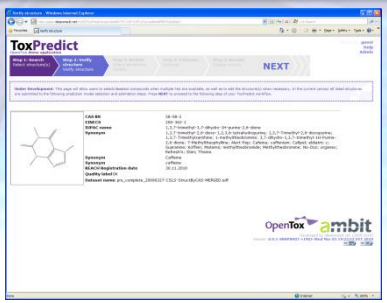
OT Dataset Service

text/uri-list,
application/rdf+xml,
chemical/x-daylight-smiles
chemical/x-mdl-sdfile,…



Here is the list of structures as URI links, RDF, MOL or SMILES.

ToxPredict: Step 2 (Verify structure(s))



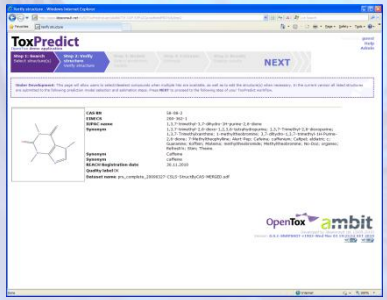
Select and/or edit structure(s)

ToxPredict Web Application

OT Dataset API ***HTTP GET***
(***HTTP POST*** for structure editing)

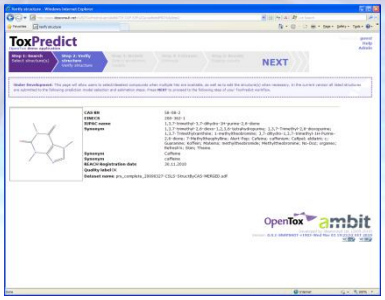
OT Dataset Service

text/uri-list,
application/rdf+xml,
chemical/x-daylight-smiles
chemical/x-mdl-sdfile,
image/png



Here is the list of structures as URI links, RDF, MOL, SMILES or images.

ToxPredict: Step 3 (Select model(s))



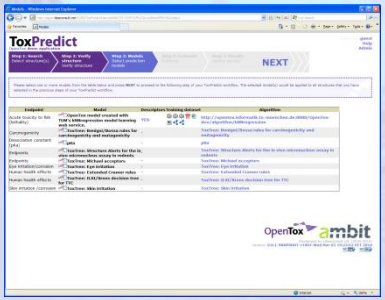
What prediction models are available? Is there a model for endpoint X?

ToxPredict Web Application

HTTP GET SPARQL query

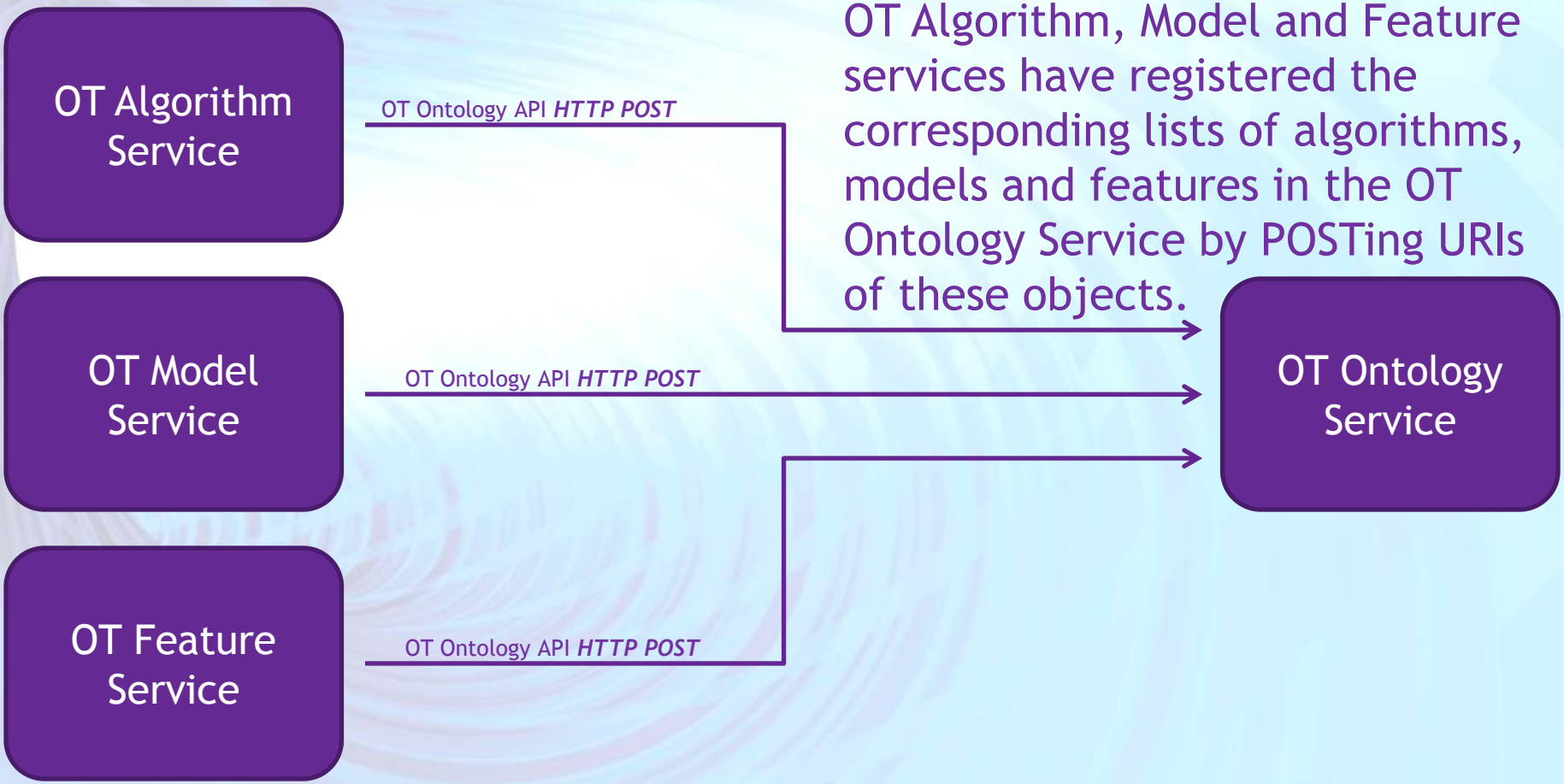
OT Ontology Service

application/sparql-results+xml



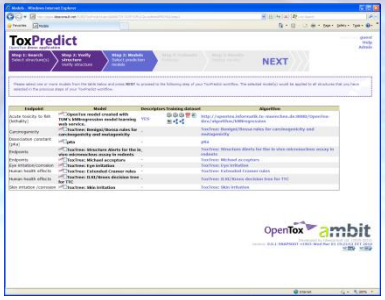
Here is the list of model URIs and related endpoints and algorithms in SPARQL format.

ToxPredict: Step 3 (behind the scenes)



OT Algorithm, Model and Feature services have registered the corresponding lists of algorithms, models and features in the OT Ontology Service by POSTing URIs of these objects.

ToxPredict: Step 4 (Estimate)



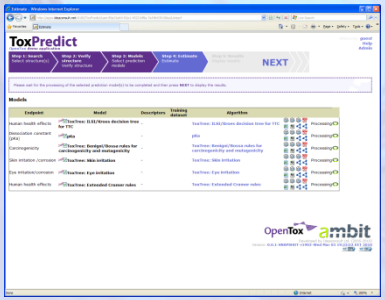
Run the selected models.

ToxPredict Web Application

OT Model API ***HTTP POST*** with parameter dataset URI from steps 1-2

OT Model Service

HTTP code 202 "Accepted"
Model task URI in HTTP:location header

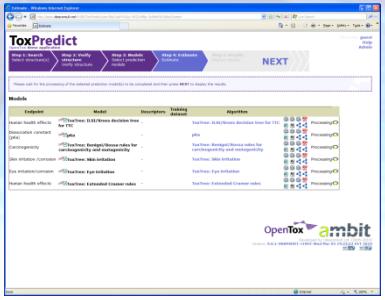


The calculation will take a while; here is a task URI, which can be queried for processing status.

Create a new task.

OT Task Service

ToxPredict: Step 4 (Estimate)



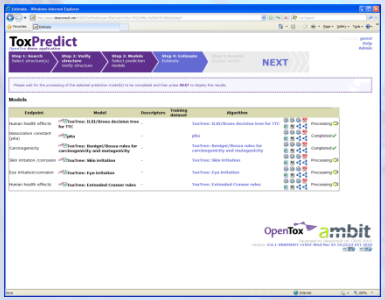
Is the task completed?

ToxPredict Web Application

OT Task API ***HTTP GET*** on task URI

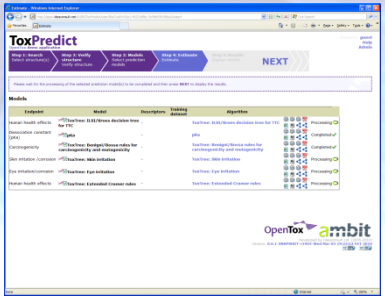
OT Task Service

HTTP code 303 "Redirect"
Dataset task URI in
HTTP:location header



Not yet, but processing has finished and the results have been posted to the OT Dataset service; here is a task URI for the dataset import.

ToxPredict: Step 4 (Estimate)



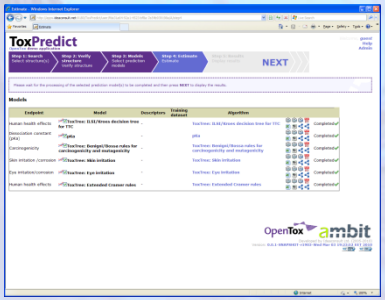
Is the task completed?

ToxPredict Web Application

OT Task API ***HTTP GET*** on task URI

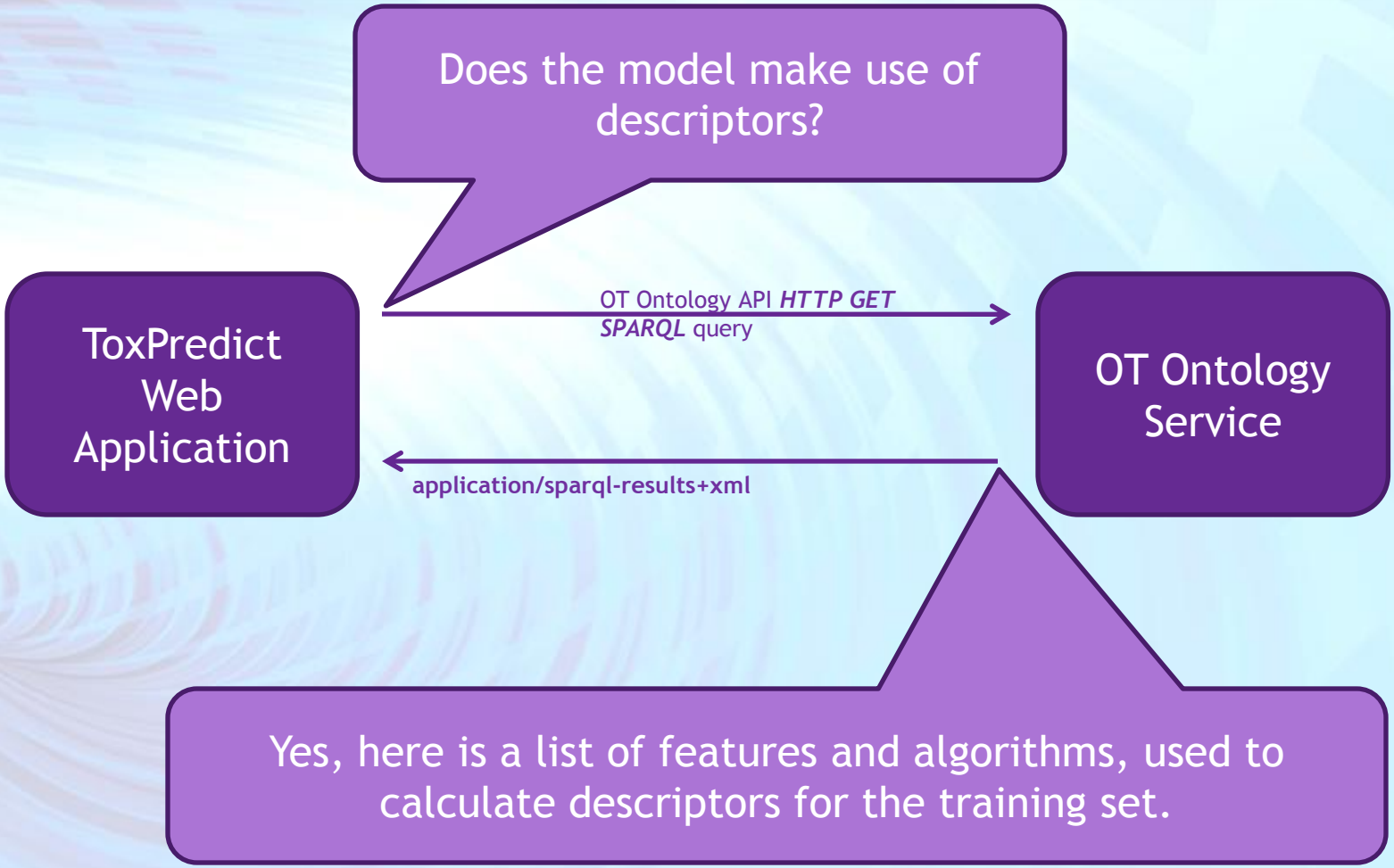
OT Task Service

HTTP code 200 "OK"
Dataset URI in
HTTP:location header

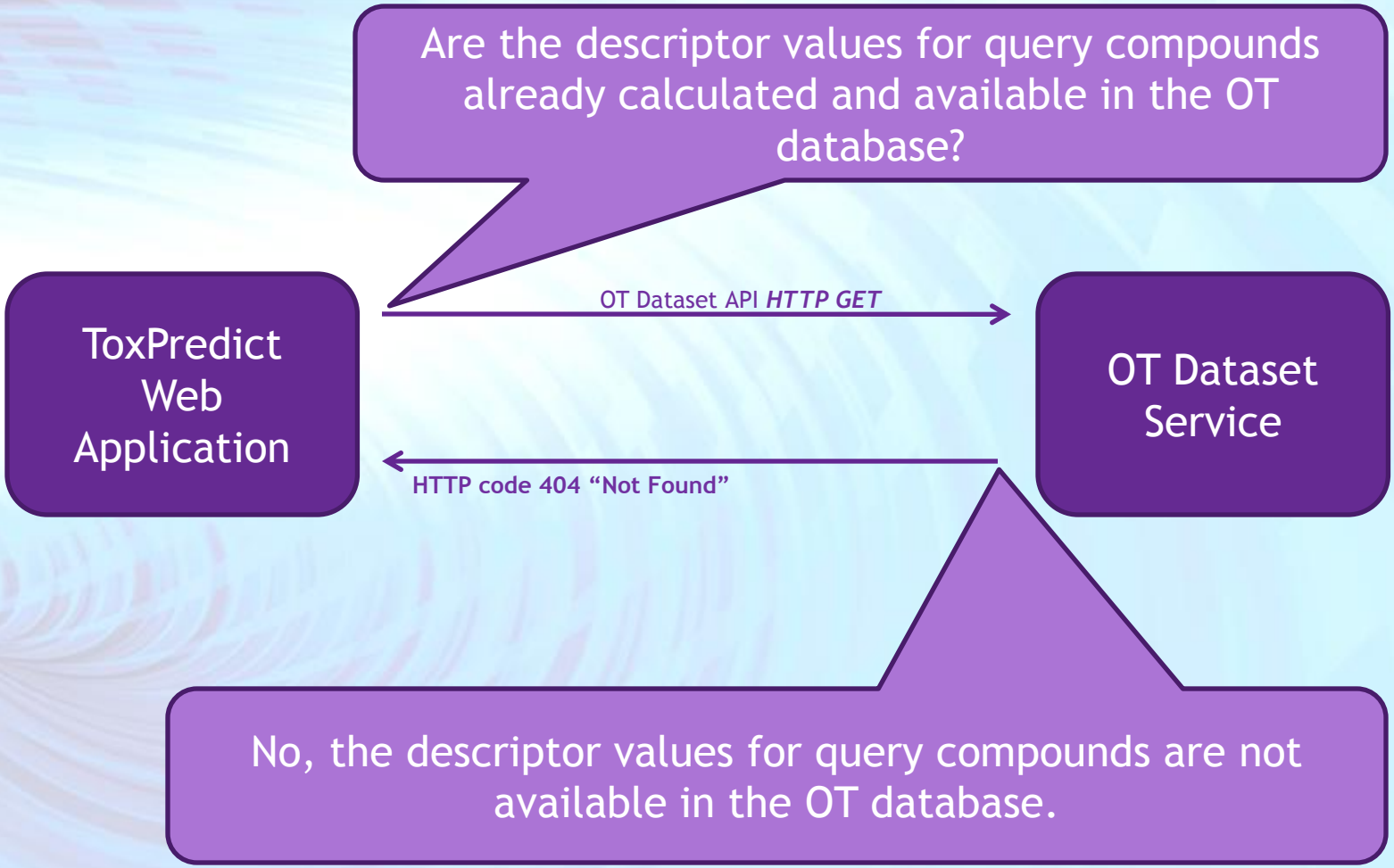


Yes, here is the dataset URI of the results.

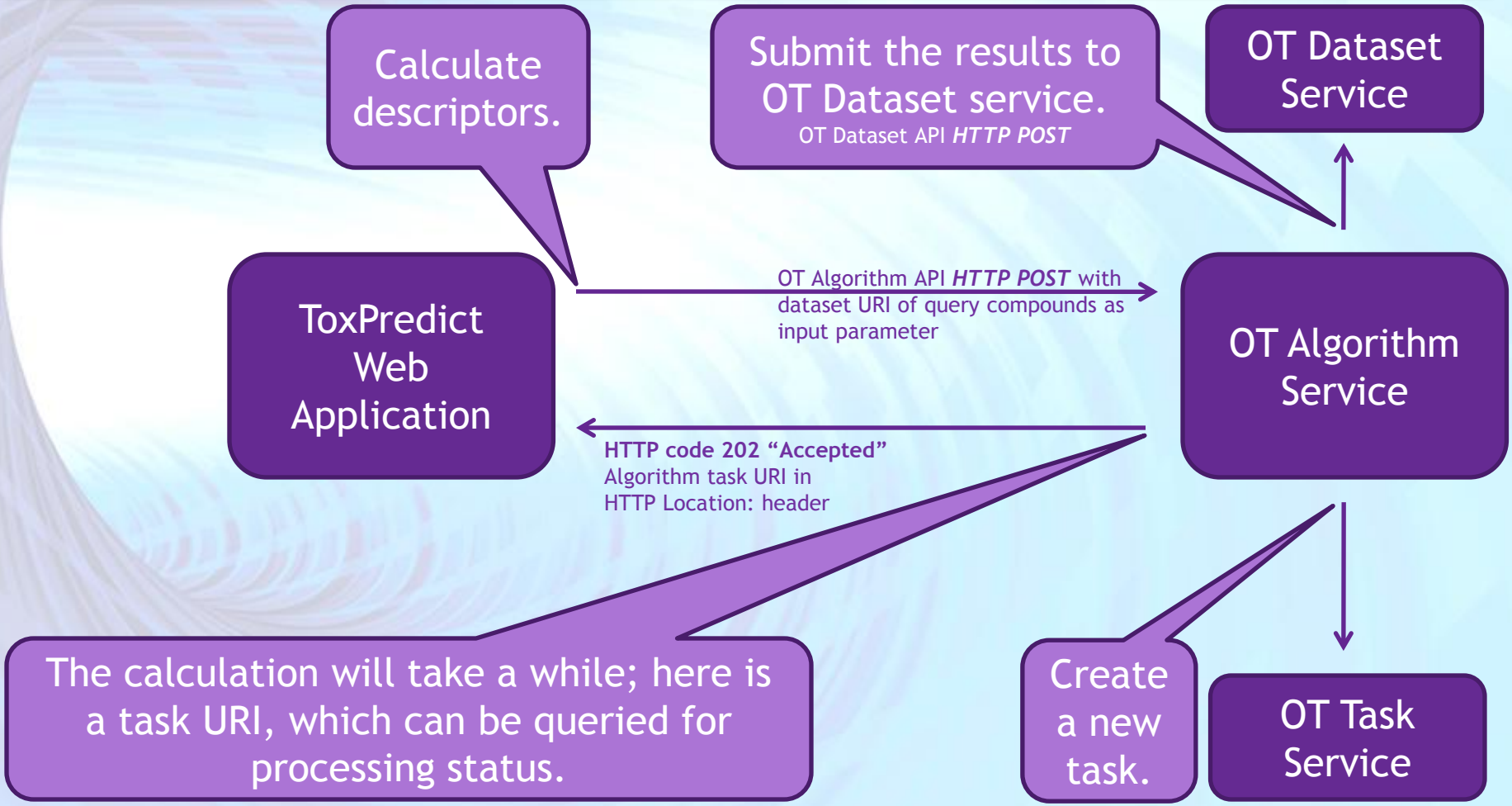
ToxPredict: Step 4 (behind the scenes)



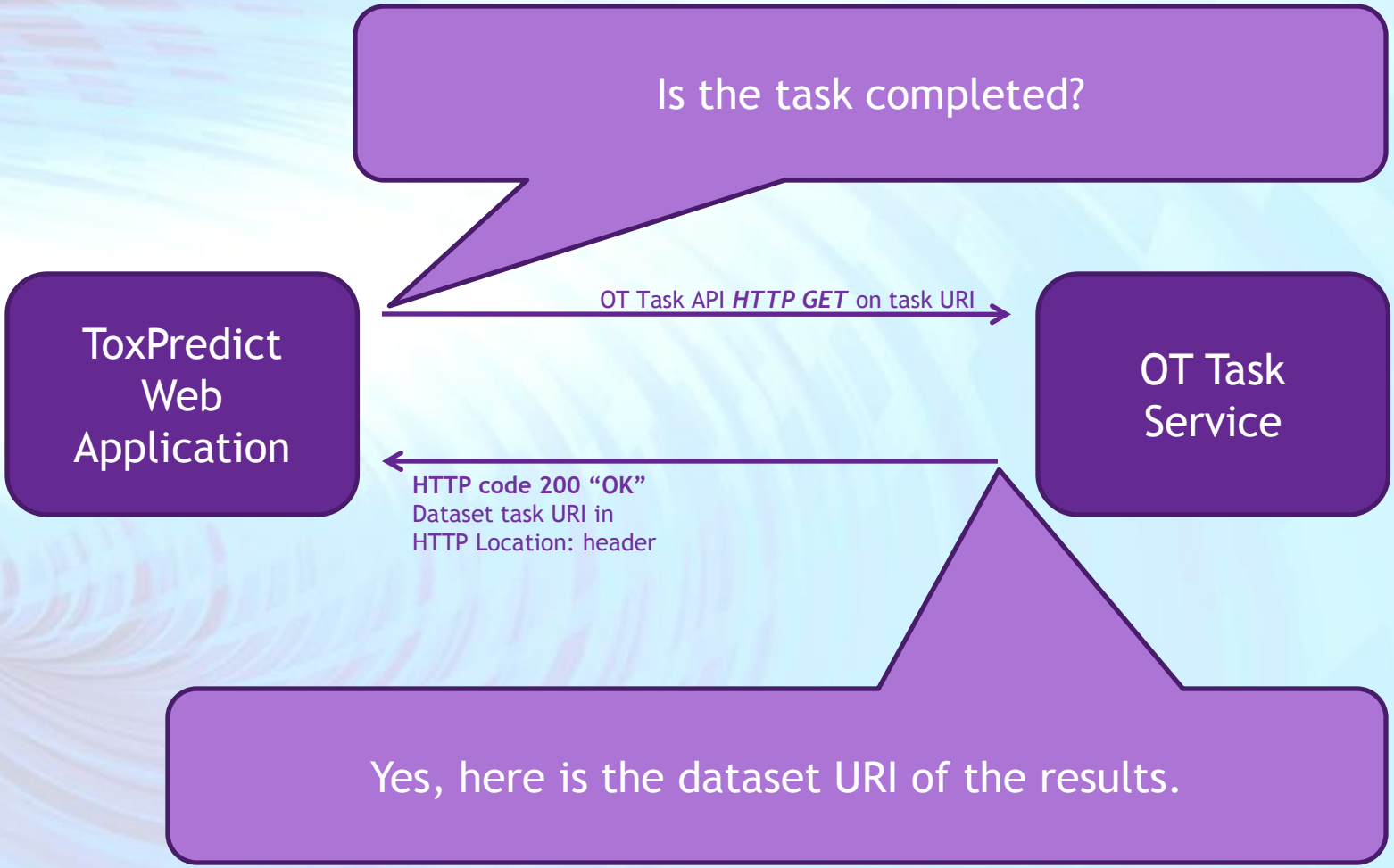
ToxPredict: Step 4 (behind the scenes)



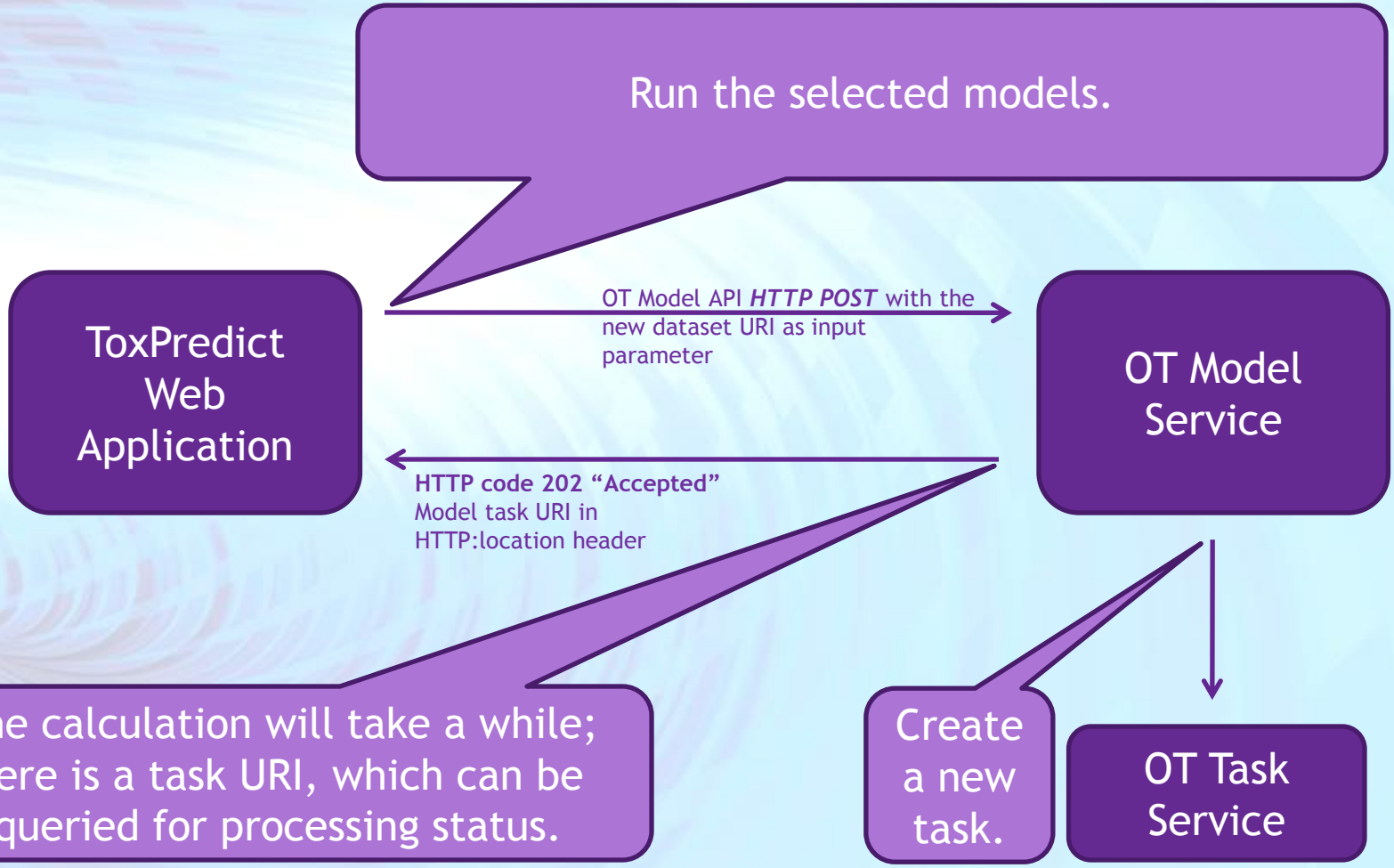
ToxPredict: Step 4 (behind the scenes)



ToxPredict: Step 4 (behind the scenes)

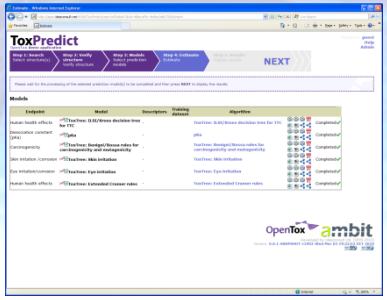


ToxPredict: Step 4 (behind the scenes)



… the remaining part of the Estimation step has been described already in slides 15 and 16

ToxPredict: Step 5 (Display results)



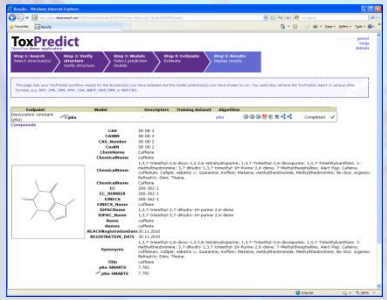
Retrieve calculation results from the final dataset URI, obtained in Step 4 (Estimation).

ToxPredict Web Application

OT Dataset API ***HTTP GET***

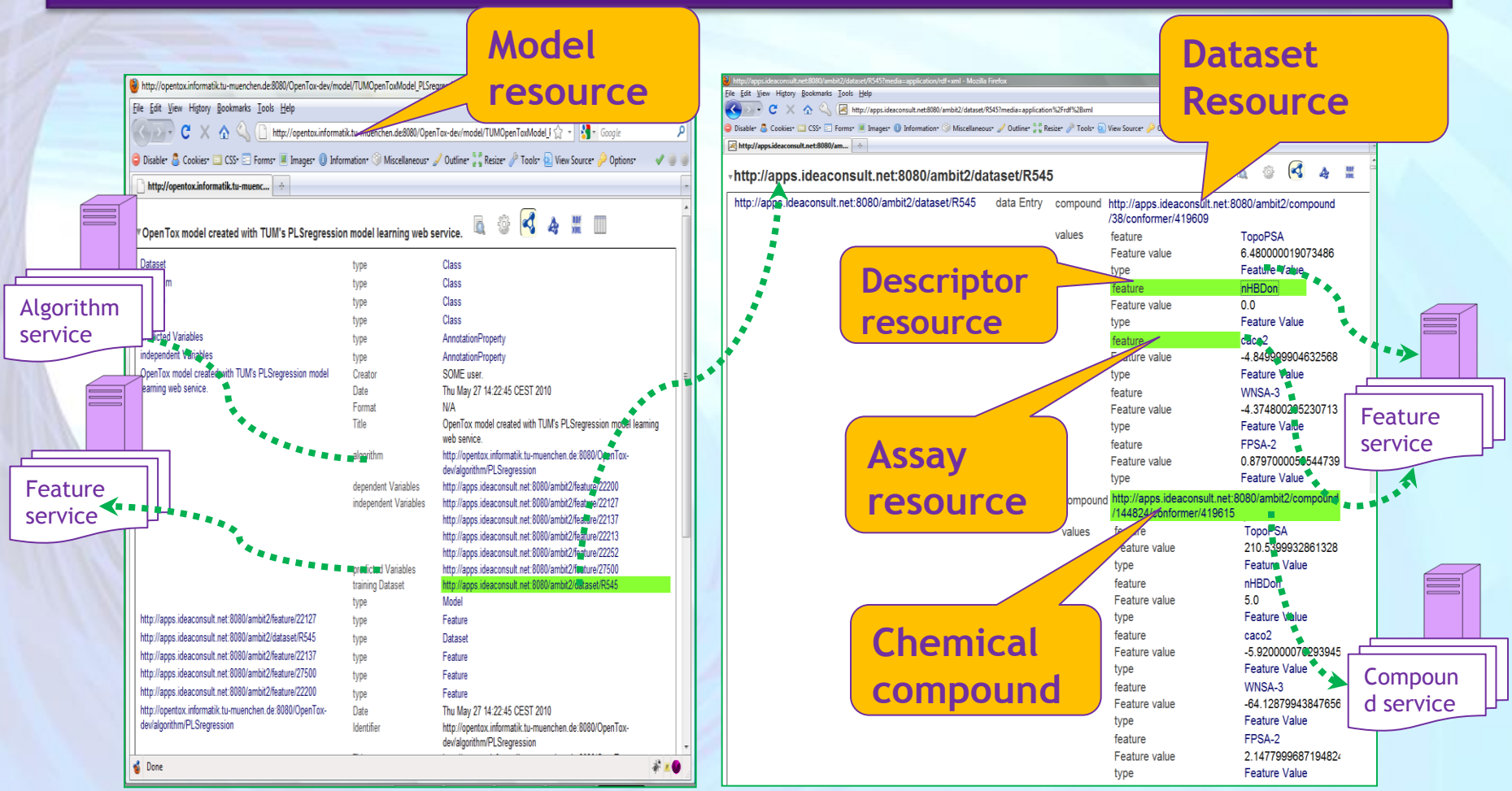
OT Dataset Service

application/rdf+xml



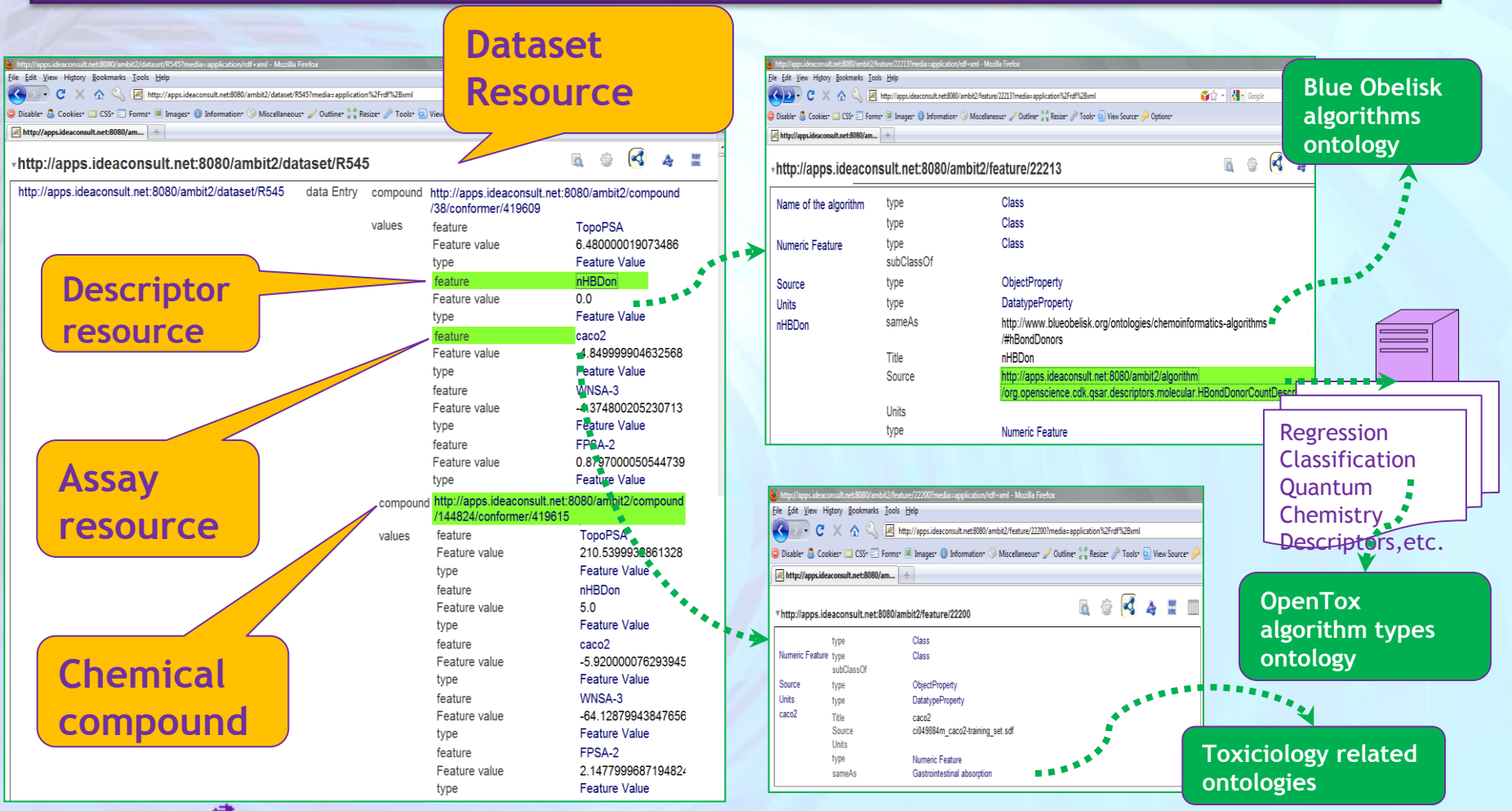
Here is the dataset content in RDF format, according to OpenTox.owl and containing estimation results, as well as compound identifiers and experimental data.

Linked resources: Compound, Algorithm, Model, Dataset, Features



OpenTox

Linked resources:
Compound, Algorithm, Model, Dataset, Features



OpenTox
Ideaconsult Ltd.